PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/54, 9/10, 15/82, A01H 1/00, 5/00, C12Q 1/68, C07K 16/40, G01N 33/53

A2

(11) International Publication Number: **WO 00/49157**

(43) International Publication Date: 24 August 2000 (24.08.00)

(21) International Application Number: PCT/US00/04542

(22) International Filing Date: 22 February 2000 (22.02.00)

(30) Priority Data:
60/121,038 22 February 1999 (22.02.99) US

(71) Applicants *(for all designated States except US)*: UNIVERSITY OF FLORIDA [US/US]; P.O. Box 115500, Gainesville, FL 32611–5500 (US). MICHIGAN STATE UNIVERSITY [US/US]; East Lansing, MI 48824–1319 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HANSON, Andrew, D. [US/US]; 507 NW 39th Road #163, Gainesville, FL 32607 (US). GAGE, Douglas, A. [US/US]; 4817 Ardmore, Okemis, MI 48864 (US).

(74) Agent: MOORE, Mark, D.; Williams, Morgan & Amerson, P.C., Suite 250, 7676 Hillmont, Houston, TX 77040 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: COMPOSITIONS AND METHODS FOR ALTERING SULFUR CONTENT IN PLANTS

(57) Abstract

Disclosed are compositions and methods for increasing the nutritional value of plants and plant parts. In illustrative embodiments S–adenosylmethionine:methionine S–methyltransferase polynucleotide and polypeptide compositions are disclosed as well as their use in modulating the levels of organic sulfur compounds, and particularly, sulfur–containing amino acids in plants and seeds derived therefrom.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

DESCRIPTION

COMPOSITIONS AND METHODS FOR ALTERING SULFUR CONTENT IN PLANTS

1.0 BACKGROUND OF THE INVENTION

The present application is a continuing application that claims priority to copending United States Provisional Patent Application Serial Number 60/121,038, filed February 22, 1999, the entire contents of which is specifically incorporated herein by reference in its entirety. The United States government has certain rights in the present invention pursuant to Grant numbers IBN-9514336, IBN-9816075, IBN-9628750 and IBN-9904263 from the National Science Foundation.

1.1 FIELD OF THE INVENTION

The present invention relates to the genetic manipulation of plants, particularly to methods and compositions for altering sulfur metabolism in plants and plant seeds, and methods for the preparation of transgenic plants that over-express the S-adenosylmethionine:methionineS-methyltransferasepolypeptide.

1.2 DESCRIPTION OF RELATED ART

Sulfur in its reduced form plays an important role in plant metabolism, being involved in the biosynthesis of a wide range of primary and secondary S-containing metabolites. In plants, sulfur metabolism includes the uptake of sulfate from the environment, assimilation into organic compounds, and channeling into proteins and secondary substances.

Plants and microorganisms are able to reduce sulfate to sulfide for synthesis of the thiol group of cysteine. Sulfate is first activated by ATP sulfurylase, forming 5'- adenylylsulfate (APS). APS can be phosphorylated by APS kinase, forming 3'- phosphoadenosine-5'-phosphosulfate (PAPS). Either APS or PAPS can be used for sulfate reduction. Generally, prokaryotes and fungi use PAPS, whereas photosynthetic eukaryotes use APS.

Cysteine, methionine, and sulfur-containing vitamins such as biotin or thiamine are essential in human nutrition. Sulfur-mediated functions include electron transport in Fe/S-clusters, structural and regulatory roles *via* protein disulfide bridges, and catalytic centers.

Additionally, secondary sulfur compounds include signaling molecules, anti-carcinogens and atmospheric compounds (Hell, 1997).

Plant methionine (Met) metabolism differs from that in other organisms by having S-methylmethionine (SMM). SMM is a ubiquitous constituent of the free amino acid pool in flowering plants, occurring in leaves, roots and other organs at levels that typically range from 0.5 to 3 μmol g^{-1} dry weight and are often higher than Met or S-adenosylmethionine (AdoMet) levels (Giovanelli *et al.*, 1980; Mudd *et al.*, 1990; Bezzubov et al., 1992). SMM has also been detected as a metabolite of radiolabeled L-Met in all flowering plants tested (Paquet *et al.*, 1995). SMM is formed from L-Met *via* the action of AdoMet:Met S-methyltransferase (MMT, EC 2.1.1.12), and can be reconverted to Met by donating a methyl group to L-homocysteine (Hcy) in a reaction catalyzed by Hcy S-methyltransferase (HMT, EC 2.1.1.10) (Giovanelli *et al.*, 1980; Mudd *et al.*, 1990) (See FIG. 1). The tandem action of MMT and HMT - together with S-adenyosylhomocysteine hydrolase - constitutes the SMM cycle, which is apparently futile (Mudd *et al.*, 1990).

A number of methods have been described for increasing sulfur amino acid content of plants. Generally, these methods provide for the overexpression of a high methionine seed storage protein. The method entails overexpressing the seed storage protein in a transformed plant. Previously, methods for increasing the sulfur amino acid content of crops were attempted through breeding. However, these methods have met with limited success. There is therefore a need for methods for raising the levels of the sulfur amino acids in plants and plant seeds.

1.3 Deficiencies in the Prior Art

Plant protein is often deficient in essential amino acids such as lysine and tryptophan, and may be particularly deficient in the sulfur amino acids, such as methionine and cysteine. As a result, diets must be supplemented with these amino acids in order to provide a balanced diet. A goal of plant breeding has been to increase the amount of sulfur amino acids present in the seed. What is lacking in the prior art are compositions and methods for preparing transgenic plants and the seeds derived therefrom that have an increased level of sulfur-containing amino acids. Also what is lacking are methods for modulating the sulfur biosynthesis of a plant cell, and methods for altering the activity of MMT polypeptides in plant cells, plant tissues, and transgenic plants, particularly, in commercially-relevant plant species that would benefit from an increased level of sulfur-containing amino acids.

2.0 SUMMARY OF THE INVENTION

The present invention overcomes these and other limitations inherent in the prior art by providing compositions and methods for increasing the nutritional value of plants. Methods are also provided particularly for modulating the levels of organic sulfur compounds, such as sulfur-containing amino acids, in plants, plant tissues, and seeds. Disclosed are novel amino acid sequences comprising S-adenosylmethionine:methionine S-methyltransferase (AdoMet:Met S-methyltransferase; MMT) polypeptides, and the polynucleotides that encode these enzymes. Also disclosed are methods of identifying MMT-specific polypeptide and polynucleotide compositions, methods for preparing recombinant host cells, vectors, virus, and expression constructs, and methods for making transgenic plants that over-express MMT-specific genes. These compositions find particular use in modulating the levels of at least one organic-sulfur compound in plants, and particularly those found in plant tissues, such as fruits and seeds.

The invention provides an isolated polynucleotide that: (a) encodes a polypeptide having MMT activity and that comprises an at least 7 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8; (b) encodes a polypeptide having MMT activity and at least about 75% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8; (c) comprises an at least 31 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or (d) hybridizes to the sequence of from position 82 to 3345 of SEQ ID NO:1 or to the sequence of from position 70 to 3282 of SEQ ID NO:3, or to the sequence of from position 1 to 610 of SEQ ID NO:7, or that hybridizes to the complement thereof, under stringent hybridization conditions.

Preferabley the isolated polynucleotide comprises a sequence region that encodes a polypeptide having an at least 7 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8, although longer contiguous sequences such as at least 9, at least 11, at least 13, at least 15, or at least 17 or more contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8 are also contemplated to be particularly preferred.

In illustrative embodiments, the isolated polynucleotide comprises a sequence region that encodes a polypeptide having the sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

The isolated polynucleotides of the invention preferably comprise a sequence region that encodes a polypeptide having MMT activity and at least about 70%, 75%, 78%, or 80% or greater sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

More preferably, the polypeptides have at least about 85%, about 90%, about 95% or about 98% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

Preferred polynucleotides of the present invention typically will comprise an at least 31 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7, although longer contiguous nucleotide sequences from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7, such as about 33, about 35, about 37, about 39, about 40, about 45, about 50, about 55 or more contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7, are also highly preferred. In fact, the MMT-encoding polynucleotides may comprise all or substantially all of the the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7.

Preferred polynucleotide compositions of the present invention typically will comprise a sequence region that hybridizes to the sequence of from position 82 to 3345 of SEQ ID NO:1 or to the sequence of from position 70 to 3282 of SEQ ID NO:3, or to the sequence of from position 1 to 610 of SEQ ID NO:7, or to the complement thereof, under stringent hybridization conditions. Such stringent hybridizations are well known to those of skill in the art, as are the methods for obtaining and identifying polynucleotides that hybridize to a selected target sequence. For example, as described hereinbelow, stringent hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C are particularly preferred.

A further embodiment of the present invention concerns an isolated polynucleotide that comprises: (a) a sequence region that consists of at least 31 contiguous nucleotides that have the same sequence as, or are complementary to, at least 31 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:7; or (b) a sequence region of from 31 to about 5000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C. Such polynucleotides may range in size from on the order of about 100 to about 10,000 nucleotides in length, with intermediate ranges such as from about 1000 to about 8,000 nucleotides in length, or from about 2000 to about 6,000 nucleotides in length, or from about 3000 to about 4000 nucleotides in length being particular preferred.

Preferred polynucleotide compositions will typically comprise an RNA, a PNA, or a DNA segment, as described hereinbelow. Such compositions may be comprised within a recombinant vector such as a plasmid, cosmid, phage, phagemid, baculovirus, bacterial artificial chromosome, or

yeast artificial chromosome vector. Likewise, the disclosed polynucleotides may be comprised within a recombinant virus or virion. It may be operably linked to a promoter, and particularly to a heterologous promoter such as a a plant-expressible constitutive, inducible, or tissue-specific promoter. Exemplary plant-expressible promoters include those such as corn sucrose synthetase 1, corn alcohol dehydrogenase 1, corn light harvesting complex, corn heat shock protein, pea small subunit RuBP carboxylase, Ti plasmid mannopine synthase, Ti plasmid nopaline synthase, petunia chalcone isomerase, bean glycine rich protein 1, Potato patatin, lectin, CaMV 35S, ALS, ubiquitin, globulin 1, cruciferin, napin, B-conglycinin, phaseolin, gama zein, or the S-E9 small subunit RuBP carboxylase promoter.

The polynucleotide compositions may encode a polypeptide that comprises the sequence of any one of SEQ ID NO:26 to SEQ ID NO:51, and may comprise the sequence of any one of SEQ ID NO:9 to SEQ ID NO:25. Such polynucleotides may be used, for example, in preparing a recombinant vector, a transgenic plant, or a recombinant polypeptide composition.

Such polynucleotide compositions may also be used as a probe for screening a plant nucleic acid library to identify a gene encoding a polypeptide having MMT activity. Alternatively, their sequence information may be used in the preparation of a target sequence probe to employ a computer-based algorithm to search a computerized database of sequences such as genomic, or expressed sequence tags, cDNAs, and the like to identify a gene encoding a polypeptide having MMT activity.

The recombinant vectors of the present invention may also be used in producing a transformed plant cell or plant tissue, a pluripotent plant cell, or a transgenic plant that expresses a polypeptide having MMT activity.

In a related embodiment, the invention provides a host cell that comprises such a recombinant vector that has at least a first heterologous expression unit comprising an MMT polynucleotide. Such a host cell may be a bacterial cell such as an *Escherichia, Salmonella* or *Agrobacterium* cell, or alternatively, may be an eukaryotic cell, such as a plant cell. Alternatively, the polynucleotide may be comprised within a virus, virion, or viral vector.

The invention also provides an isolated polypeptide encoded by the disclosed MMT polynucleotides. Such polypeptides preferably comprise an at least 7, 8, 9, 10, 11, 12 13, 14, 15, 16, 17, 18, 19, or 20 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8, and preferably share at least about 65%, about 70%, about 75%, about 80%, about 85%,

about 90%, about 95%, or about 99% or higher sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

Such polypeptides may be used in the preparation of an antibody that specifically binds to an MTT polypeptide, using the immunological methods described hereinbelow. The MMT antibody compositions so produced that specifically binds to the disclosed MMT polypeptides also represent an important embodiment of the present invention. Such antibodies may be suitably packaged in an immunodetection kit, along with an immunodetection reagent, and instructions for using the antibody in methods such as ELISAs and other immunoaffinity methodologies to detect the presence of MMT polypeptides in a target sample, such as in a plant or leaf extract.

The invention also provides nucleic acid detection kits that typically comprise in suitable container means, at least a first isolated nucleic acid segment comprising an MMT polynucleotide, a detection reagent, and instructions for using the MMT-specific nucleic acid segment to detect other MMT sequences or to use as probes or primers for related and DNA sequencing methodologies and the like.

Compositions are also provided by the invention that comprise: (a) an MMT polynucleotide, (b) an MMT polypeptide, (c) an MMT-specific antibody, (d) a recombinant vector, (e) virus, or (f) host cell that expresses an MMT polynucleotide or polypeptide.

The invention further provides a transgenic plant that comprises: (a) a heterologous nucleic acid segment that comprises an MMT polynucleotide; (b) a transformed host cell that expresses an MMT polypeptide; (c) a recombinant virus that expresses an MMT polypeptide; or (d) a recombinant vector that encodes an MMT polypeptide.

The transgenic plant preferably has stably incorporated into its genome a heterologous nucleic acid segment that comprises an MMT polynucleotide, wherein the polynucleotide is operably linked to a promoter that expresses the polynucleotide in the cells and tissues of the transgenic plant. Such transgenic plants are preferably monocotyledonous or dicotyledonous plants, such as grains, trees, legumes, fibers, vegetables, fruits, berries, nuts, citrus, grasses, cacti, succulents, flowers, or other ornamental plants.

Exemplary plants include, but are not limited to, corn, rice, millet, tobacco, alfalfa, soybean, bean, sorghum, pea, *Brassica*, safflower, potato, coconut, palm, pumpkin, squash, poppy, sesame, peanut, cocoa, coffee, tomato, flax, canola, sunflower, cotton, flax, kapok, wheat, oat, barley, walnut, pecan, almond, and rye.

The invention further discloses and claims progeny of any generation of such transgenic plant, as well as the seed of any generation of such transgenic plants, and seed of any generation, offspring, or subsequent progeny of such transgenic plants. Particularly encompassed by the invention are seeds, nuts, legumes, and the like, that have an increased level of sulfur-containing amino acids, relative to untransformed plants of the same species that do not contain one or more exogenously provided MMT-encoding transgenes. Such seeds are particularly preferred for animal foodstuffs, as well as those having increased protein and nutrition content suitable for human consumption.

The invention also provides hereinbelow methods for detecting an MMT-encoding polynucleotide in a sample. Such a method typically involves the steps of: (a) contacting a population of polynucleotides suspected of encoding an MMT polypeptide with at least a first labeled MMT polynucleotide, under conditions effective to allow hybridization of substantially complementary nucleic acids; and (b) detecting the hybridized complementary nucleic acids so formed.

A method is also provided for detecting an MMT polypeptide in a biological sample. This method typically involves contacting a biological sample suspected of containing an MMT polypeptide with a labeled MMT-specific antibody, under conditions effective to allow the formation of immune complexes, and detecting the immune complexes so formed.

A method is also provided for increasing the amount of an MMT polypeptide in a plant cell. This method typically involves expressing in such a plant cell a biologically effective amount of an MMT polynucleotide. The term "biologically effective amount" will be understood by the skilled artisan to mean an amount of the polynucleotide composition that is effective to produce the desired phenotypic trait in the resulting transformed plant cell, *i.e.*, an increased level or amount of MMT polypeptide or MMT enzymatic activity in the cell when compared to a similar untransformed or "wild-type" plant cell.

A method of increasing the level of sulfur amino acids in a plant cell is also provided by the invention. This method generally comprises, expressing in a suitable plant cell a biologically effective amount of an MMT polynucleotide. Such an effective amount will be recognized by the skilled artisan as an amount necessary to alter, increase, or improve the level or extent of sulfur containing amino acids in the plant cells and plant tissues. This method is particularly useful for increasing the level of cysteine or methionine or other sulfur-containing compounds in the cells, fruits, seeds, and tissues of the transformed plant.

This method provides means for modulating the biosynthesis of a sulfur compound in a plant, increasing the yield of plant protein, or improving the overall nutritional value of the plant for consumption by animals and/or humans.

An exemplary method may further comprise the steps of growing the transformed plant under conditions effective for obtaining seeds from the plant, and collecting the seeds so produced by the transformed plant. This method is particularly desirable for the recovery of high-protein, high-nutritional value seeds, grains, nuts, flours, and the like. The method may also further comprise the step of (e) transforming the plant cell with one or more additional polynucleotides that encode one or more sulfur compound synthesis enzymes, such as one or more enzymes in the pathway for methionine synthesis, to further increase in the cells and tissues of the plant the level of sulfur-containing amino acids. Increasing the amount of sulfur-containing amino acids in plant seed, may be achieved by further growing such a transgenic plant, under conditions effective to produce seed and then obtaining the seed produced from the plant.

2.1 POLYPEPTIDE COMPOSITIONS

In a first embodiment, the invention provides polypeptides, peptides and proteins that comprise all, substantially all, or portions of an MMT enzyme. Highly preferred MMT polypeptides are those that comprise an at least about 15, an at least about 16, an at least about 17, an at least about 18, an at least about 19, or an at least about 20 or more contiguous amino acid sequence from one of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, and that have MMT enzymatic activity when expressed in a suitable plant host cell cultured under the appropriate conditions for MMT expression and enzymatic activity. Likewise, MMT polypeptides that comprise an at least about 21 or 22 or 23 or 24 contiguous amino acid sequence from one of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, and that have MMT enzymatic activity when expressed in a suitable plant host cell are also contemplated to be particularly useful in the methods disclosed herein.

In certain circumstances, it may be desirable to employ MMT polypeptides that are even more homologous to the sequences disclosed in SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8. In those embodiments, the MMT polypeptides of the invention will preferably comprise a primary amino acid sequence that comprises an at least about 25, and at least about 30, an at least about 35, an at least about 40, an at least about 45, or an at least about 50 or so

contiguous amino acid sequence selected from one or more of the sequences of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. Likewise, in other embodiments, it may be desirable to employ MMT polypeptides that are even more homologous to the sequences disclosed in SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8. In those embodiments, the MMT polypeptides of the invention will preferably comprise a primary amino acid sequence that comprises an at least about 55, and at least about 60, an at least about 65, an at least about 70, an at least about 75, or an at least about 80 or so contiguous amino acid sequence selected from one or more of the sequences of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. In fact, when more highly homologous MMT polypeptides are contemplated, those having an at least about 85, and at least about 90, an at least about 95, an at least about 100, an at least about 110, an at least about 120, an at least about 125, an at least about 130, an at least about 135, an at least about 140, an at least about 145, or an at least about 150, or so contiguous amino acid sequence selected from one or more of the sequences of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8 will be particularly preferred.

Shorter peptide and polypeptide sequences comprised with one or more of the disclosed MMT proteins are also within the scope of the present invention. Such peptides may be utilized as described herein in the preparation of epitopes, or used as antigens for the generation of MMT-specific antibodies, or may even be used to screen antibody samples for species that specifically bind to an MMT peptide motif. Such smaller peptides include, but are not limited to the sequences set forth in SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41 SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, and SEQ ID NO:51 are particularly useful as probes for identifying polypeptides of the MMT family that share conserved regions. These sequences, which are conserved within two or more of the disclosed MMT polypeptides, are illustrated in FIG. 6A, FIG. 6B, FIG. 6C, FIG. 6D, FIG. 8A and FIG. 8B.

2.2 POLYNUCLEOTIDE COMPOSITIONS

In a second embodiment, the invention concerns polynucleotides that encode the MMT polypeptides of the present invention. Such sequences preferably comprise from at least about 20, to at least about 3300 or more contiguous nucleotides from any one of SEQ ID NO:1, SEQ

ID NO:3, SEQ ID NO:5, or SEQ ID NO:7. As such, polynucleotides that comprise at least about 30 to about 2500 or more contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 are contemplated to be particularly preferred in the methods of the present invention. Similarly, polynucleotides that comprise at least about 40 to about 1500 or more contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 are also contemplated to be particularly preferred in the methods of the present invention, as are those polynucleotides that comprise at least about 50 to about 1000 or more contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7, and those polynucleotides that comprise at least about 60 to about 500 or more contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7. Naturally, all intermediate contiguous sequences are contemplated to fall within the scope of the present invention. For example, polynucleotides that comprise at least about 31, at least about 32, at least about 33, at least about 34, at least about 35, at least about 36, at least about 37, at least about 38, at least about 39, at least about 40, at least about 41, at least about 42, at least about 43, at least about 44, at least about 45, at least about 46, at least about 47, at least about 48, at least about 49, at least about 50, at least about 51, at least about 52, at least about 53, at least about 54, at least about 55, at least about 56, at least about 57, at least about 58, at least about 59, at least about 60, at least about 61, at least about 62, at least about 63, at least about 64, at least about 65, at least about 66, at least about 67, at least about 68, at least about 69, at least about 70, at least about 71, at least about 72, at least about 73, at least about 74, at least about 75, at least about 76, at least about 77, at least about 78, at least about 79, at least about 80, at least about 85, at least about 90, at least about 95, or at least about 100 or more contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 are contemplated to be particularly preferred in the methods of the present invention, and are contemplated to be particularly preferred polynucleotide compositions.

Likewise, MMT-encoding polynucleotides that comprise at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190, or at least about 200 or more contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 are contemplated to be particularly preferred polynucleotide compositions. MMT-specific polynucleotides that comprise at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, or more contiguous

nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7 are contemplated to be particularly preferred polynucleotide compositions.

When it is desirable to employ MMT-encoding polynucleotides that are significantly more homologous to the polynucleotide sequences of either the *Arabidopsis*, *Zea*, *Nicotiana*, or *Wollastonia* sequences disclosed herein, polynucleotide compositions may be selected that encode MMT or MMT-derived peptides that comprise at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, at least about 900, at least about 1000, at least about 1100, at least about 1200, at least about 1300, at least about 1400, at least about 1500, at least about 1600, at least about 1700, at least about 1800, at least about 1900, at least about 2000, at least about 2100, at least about 2200, at least about 2300, or more contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5, are contemplated to be particularly preferred polynucleotide compositions, including those sequences that comprise at least about 2400, about 2500, about 2600, about 2700, about 2800, about 2900, about 3000, about 3100, and even those up to and including the full-length DNA sequences disclosed in SEQ ID NO:1, SEQ ID NO:3, and SEQ ID NO:5.

Likewise, the MMT polynucleotide compositions of the present invention also encompass those nucleic acid segments that encode a polypeptide having MMT activity, and that comprise a nucleic acid sequence of at least about 7 or 8 contiguous amino acids from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8. In more preferred embodiments, the MMT polypeptides of the present invention comprise an at least about 9 or 10 contiguous amino acid sequence from one of these full-length sequences. When it is desirable to identify MMT polypeptides that are still more homologous to the disclosed *Wollastonia*, *Arapidopsis*, *Zea*, or *Nicotiana* sequences, one may wish to utilize MMT polypeptides that comprise an at least about 11, about 12, about 13, or about 14 or more contiguous amino acid sequence from one of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.

2.3 Nucleic Acid Probes and Primers for MMT Gene Detection, Amplification, and Sequencing

In related embodiments, the invention provides methods and compositions for detecting homologous MMT-encoding polynucleotides and homologous MMT polypeptides.

For detection and sequencing of polynucleotides, it is often desirable to isolate smaller polynucleotides for use as hybridization probes, synthesis or sequencing primers, and the like as

described in detail herein. In such embodiments, shorter polynucleotide sequences are particularly desirable, including those that comprise a sequence of at least about 30 or 40 or 50 or so contiguous nucleotides from one or more of the DNA sequences disclosed in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO:7. These sequences find particular utility as probes for screening clone banks, colony blots, or as computer homology search strings for identifying homologous polynucleotide sequences *via* computer-based algorithm homology searches. This is particularly important when it is desirable to screen a database of cDNA sequences, expressed sequence tags (ESTs) or genomic or chromosomal sequence databases.

For example, polynucleotides that comprise at least about 12, at least about 13, at least about 14, at least about 15, at least about 16, at least about 17, at least about 18, at least about 19, at least about 20, at least about 21, at least about 22, at least about 23, at least about 24, at least about 25, at least about 26, at least about 27, at least about 28, at least about 29, at least about 30, at least about 31, at least about 32, at least about 33, at least about 34, at least about 35, at least about 36, at least about 37, at least about 38, at least about 39, at least about 40, at least about 41, at least about 42, at least about 43, at least about 44, at least about 45, at least about 46, at least about 47, at least about 48, at least about 49, at least about 50, at least about 51, at least about 52, at least about 53, at least about 54, at least about 55, at least about 56, at least about 57, at least about 58, at least about 59, or even at least about 60 or so nucleotides from one of the disclosed sequences are particularly suited for these embodiments. For example, the inventors have identified several sequences within the disclosed polynucleotides that are identical in at least two or more plant MMT gene sequences. These sequences include those described in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:21 and exemplified in FIG. 5A, FIG. 5B, FIG. 5C, FIG. 5D, FIG. 5E, FIG. 5F, FIG. 7A, and FIG. 7B.

2.4 Identification of MMT Homologs

In addition to the particular illustrative polypeptide and polynucleotide sequences disclosed herein, those having benefit of the present teaching are now also able to identify and characterize a wide variety of MMT-homologs and/or isozymes, as well as to identify, characterize, and sequence a variety of MMT-encoding polynucleotides from a variety of plant species. In fact, the inventors contemplate that any plant-derived MMT protein or peptide can be identified using the methods

disclosed herein and may be obtained by using the immunological methods disclosed herein to obtain MMT proteins and peptides from a variety of disparate species. Alternatively, the inventors contemplate that those of skill in the art having the benefit of the teachings disclosed herein will be able to identify MMT-encoding polynucleotides either by comparison of one or more of the disclosed sequences to computer databases of plant EST sequences, and identification of highly homologous sequences, or alternatively, by traditional hybridization screening methods employing one or more labeled MMT-specific polynucleotide sequences to screen a population of target nucleic acids, such as *e.g.*, a cDNA or other such genetic library, a colony or clone bank, or by screening individual isolates from particular plant species.

Because the inventors have successfully demonstrated the presence of MMT-specific polynucleotides and polypeptides in at least four disparate plant genera (*e.g.*, *Wollastonia*, *Arabidopsis*, *Zea*, and *Nicotiana*) using the *Wollastonia*-derived sequences, they contemplate the additional identification of related MMT polypeptides and the gene sequences that encode them. In particular, the inventors contemplate the identification of MMT variants, homologs, and related sequences using one or more of the methods disclosed herein to identify a family of MMT sequences. Likewise, one of skill in the art will even be able to utilize the teachings of the present disclosure to identify other MMT-like polypeptides and polynucleotides, including those from related and from distantly-related plant species and to use these sequences in the preparation of transgenic plants having modified sulfur biosynthesis. By hybridization, immunological, and computer-based homology algorithms, the inventors further contemplate the identification and characterization of MMT-specific compositions from species that are not yet even described or characterized as possessing MMT activity.

In addition to the particular full-length MMT polypeptides disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8, the inventors also contemplate the preparation and use of substantially full-length sequences in certain embodiments. As such, polypeptides may be obtained that comprise from at least about 80% or so, and up to and including those having at least about 99% of the full-length MMT primary amino acid sequence as disclosed herein, and yet still possess significant MMT enzymatic activity *in vitro* and *in vivo*. In fact, "truncated" polypeptides or "near-full-length" or "substantially full-length" polypeptides are well known in the plant molecular biological arts to often possess all, or almost all of the enzymatic activity that the full-length polypeptide possesses. In many embodiments, these slightly shorter polypeptide sequences may be desirable for use in many of the disclosed methods. This is particularly true,

when the creation of "chimeric" polypeptides is desired, as well as in the creation of hybrid polypeptides that have, for example, the addition of a particular amino acid sequence to "target" the localization of the polypeptide to a particular cellular location, or to a particular region of the plant in which the polypeptide is expressed. For example, the preparation of a fusion protein that possesses both MMT activity, yet further comprises a sequence region that targets the peptide to a particular cellular region, such as the membrane, or to a particular organelle, *etc.* is often desirable. As such, truncated or fusion proteins that comprise only about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% of the primary amino acid sequence as disclosed in any one of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8 are particularly embodied by the present invention.

In the same that an MMT polypeptide need not include the entire full-length sequences as disclosed in the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8 to be useful in the practice of the present methods, the primary amino acid sequence of a particular MMT polypeptide need not have complete sequence identity to one or more of the polypeptides disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8. In fact, the primary amino acid sequence of a particular MMT polypeptide need only comprise enough of the primary MMT sequence to substantially perform its enzymatic function *in vivo* or *in situ.* Owing to the nature of conservative amino acid replacement, indeed several amino acids may be changed, altered, mutagenized, or even deleted in the primary amino acid sequence of a particular MMT protein and yet still give rise to a functional MMT polypeptide which still possesses an enzymatic activity similar or identical to that of the native full-length MMTs disclosed in the accompanying examples. In fact, it is well known in the plant molecular biological arts that two polypeptides from different species may differ slightly, or even sometimes, substantially in their primary amino acid sequence, and yet, still possess the same biological activity. As such, homologous or "cognate" MMT polypeptides may be designed synthetically, site-specifically modified, or isolated from different biological sources, that possess similar MMT enzymatic activity, but yet share less than 100% identity at the primary amino acid level with one of the MMT sequences disclosed herein. In fact, such MMT homologous polypeptides may share approximately 60% or 65% sequence identity with one or more of the disclosed sequences herein. More homologous MMT sequences will include those polypeptides that are from about 70% to about 80% identical to one or more of the polypeptides of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8. Still

more homologous MMT sequences will include those polypeptides that share from about 85% to about 95% sequence identity with one or more of the polypeptides disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. When highly homologous polypeptide are identified that possess MMT enzymatic activity, such as is often the case when polypeptides are obtained from closely-related species, cultivars, or hybrids, the MMT polypeptides identified may share about 96%, about 97%, about 98%, or even about 99% or more sequence identity with one or more of the sequences disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. Naturally, all intermediate % identity values are contemplated to fall within the scope of the present disclosure. As such, polypeptides having about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, and those having about 98% primary amino acid sequence identity to one or more of the sequences disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8 are contemplated to be useful in the formulation of the methods and compositions of the present invention.

In preferred embodiments, the MMT polypeptides of the present invention comprise an amino acid sequence having at least about 7 or 8 contiguous amino acids selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8. Likewise, the MMT compositions of the present invention also encompass those polypeptides that have MMT activity, and that comprise an amino acid sequence of at least about 9 or 10 contiguous amino acids from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8. In more preferred embodiments, the MMT polypeptides of the present invention comprise an at least about 11 or 12 contiguous amino acid sequence from one of these full-length sequences. When it is desirable to identify MMT polypeptides that are still more homologous to the disclosed *Wollastonia*, *Arabidopsis*, *Zea*, or *Nicotiana* sequences, one may wish to utilize MMT polypeptides that comprise an at least about 13 or 14 or 15 or 16 contiguous amino acid sequence from one of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.

Highly preferred MMT polypeptides are those that comprise an at least about 17 or 18 or 19 or 20 contiguous amino acid sequence from one of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, and that have MMT enzymatic activity when expressed in a suitable plant host cell cultured under the appropriate conditions for MMT expression and enzymatic activity. Likewise, MMT polypeptides that comprise an at least about 21 or 22 or 23 or 24

contiguous amino acid sequence from one of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, and that have MMT enzymatic activity when expressed in a suitable plant host cell are also contemplated to be particularly useful in the methods disclosed herein.

In certain circumstances, it may be desirable to employ MMT polypeptides that are even more homologous to the sequences disclosed in SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8. In those embodiments, the MMT polypeptides of the invention will preferably comprise a primary amino acid sequence that comprises an at least about 25, and at least about 30, an at least about 35, an at least about 40, an at least about 45, or an at least about 50 or so contiguous amino acid sequence selected from one or more of the sequences of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. Likewise, in other embodiments, it may be desirable to employ MMT polypeptides that are even more homologous to the sequences disclosed in SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8. In those embodiments, the MMT polypeptides of the invention will preferably comprise a primary amino acid sequence that comprises an at least about 55, and at least about 60, an at least about 65, an at least about 70, an at least about 75, or an at least about 80 or so contiguous amino acid sequence selected from one or more of the sequences of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. In fact, when more highly homologous MMT polypeptides are contemplated, those having an at least about 85, and at least about 90, an at least about 95, an at least about 100, an at least about 110, an at least about 120, an at least about 125, an at least about 130, an at least about 135, an at least about 140, an at least about 145, or an at least about 150, or so contiguous amino acid sequence selected from one or more of the sequences of SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8 will be particularly preferred.

2.5 Recombinant Vectors

One important embodiment of the invention is a recombinant vector that comprises a nucleic acid segment encoding one or more of the novel polypeptides disclosed herein. Such a vector may be transferred to and replicated in a prokaryotic or eukaryotic host, with bacterial cells being particularly preferred as prokaryotic hosts, and plant cells being particularly preferred as eukaryotic hosts.

In preferred embodiments, the recombinant vector comprises a nucleic acid segment encoding the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, or a nucleic acid segment that encodes a polypeptide comprising one or more sequences

selected from the group consisting of SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, and SEQ ID NO:51.

Highly preferred nucleic acid segments are those which comprise an at least 31 basepair contiguous sequence from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7, or a nucleic acid segment that comprises one or more sequences selected from the group consisting of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, and SEQ ID NO:25.

Another important embodiment of the invention is a transformed host cell that expresses one or more of these recombinant vectors. The host cell may be either prokaryotic or eukaryotic, and particularly preferred host cells are those that express the nucleic acid segment(s) comprising the recombinant vector that encodes one or more MMT polypeptides. Bacterial cells are particularly preferred as prokaryotic hosts, and plant cells are particularly preferred as eukaryotic hosts

In accordance with the present invention, nucleic acid sequences include and are not limited to DNA, including and not limited to cDNA and genomic DNA, genes; RNA, including and not limited to mRNA and tRNA; PNAs (peptide nucleic acids), antisense sequences, nucleosides, and suitable nucleic acid sequences such as those set forth herein, as well as variants in the nucleic acid sequences such as alterations, deletions, mutations, and homologs capable of expressing the MMT polypeptides of the present invention.

As such the present invention also concerns DNA segments, that are free from total genomic DNA and that encode the novel MMT proteins disclosed herein. DNA segments encoding these peptide species may prove to encode proteins, polypeptides, subunits, functional domains, and the like of MMT-related or other non-related gene products. In addition these DNA segments may be synthesized entirely *in vitro* using methods that are well known to those of skill in the art.

As used herein, the term "DNA segment" refers to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a MMT

polypeptide or peptide refers to a DNA segment that contains MMT polypeptide coding sequences yet is isolated away from, or purified free from, total genomic DNA of the species from which the DNA segment is obtained. Included within the term "DNA segment", are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phagemids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified nucleic acid or gene sequence that encodes an MMT polypeptide refers to a DNA segment which may include in addition to peptide encoding sequences, certain other elements such as, regulatory sequences, isolated substantially away from other naturally occurring genes or protein-encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein-, polypeptide- or peptide-encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, operon sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides or peptides.

"Isolated substantially away from other coding sequences" means that the gene of interest, in this case, a gene encoding an MMT polypeptide, forms the significant part of the coding region of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or operon coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes, recombinant genes, synthetic linkers, or coding regions later added to the segment by the hand of man.

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that encode an MMT peptide or polypeptide species that includes within its amino acid sequence an amino acid sequence essentially as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51.

The term "a sequence essentially as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID

NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51" means that the sequence substantially corresponds to a portion of the sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51, and has relatively few amino acids that are not identical to, or a biologically functional equivalent of, the amino acids of any of these sequences. The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein (*e.g.*, see Illustrative Embodiments).

Accordingly, sequences that have between about 65% and about 75% or between about 75% and about 85%, or more preferably between about 86% and about 90%, or even more preferably between about 91% or 92% or 93% and about 97% or 98% or 99% amino acid sequence identity or functional equivalence to the amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51 will be sequences that are "essentially as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51."

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5′ or 3′ sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly

applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5′ or 3′ portions of the coding region or may include various internal sequences, *i.e.* introns, which are known to occur within genes.

The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

For example, nucleic acid fragments may be prepared that include a short contiguous stretch encoding the peptide sequence disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51, or that are identical to or complementary to DNA sequences which encode the peptide disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51, and particularly the DNA segments disclosed in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:25.

Highly preferred nucleic acid segments of the present invention comprise one or more MMT-encoding genes of the invention, or a portion of one or more MMT-encoding genes of the invention. For certain application, relatively small contiguous nucleic acid sequences are preferable, such as those which are about 14 or 15 or 16 or 17 or 18 or 19, or 20, or 21, or 22, or 23, or 24, or 25, or 26, or 27, or 28, or 29, or 30-50, 51-80, 81-100 or so nucleotides in length.

Alternatively, in some embodiments, and particularly those involving preparation of recombinant vectors, transformation of suitable host cells, and preparation of transgenic plant cell, longer nucleic acid segments are preferred, particularly those that include the entire coding region of one or more MMT-encoding genes. As such, the preferred segments may include those that are up to about 20,000 or so nucleotides in length, or alternatively, shorter sequences such as those about 19,000, about 18,000, about 17,000, about 16,000, about 15,000, about 14,000, about 13,000, about 12,000, 11,000, about 10,000, about 9,000, about 8,000, about 7,000, about 6,000, about 5,000, about 4,500, about 4,000, about 3,500, about 3,000, about 2,500, about 2,000, about 1,500, about 1,000, about 500, or about 200 or so base pairs in length. Of course, these numbers are not intended to be exclusionary of all possible intermediate lengths in the range of from about 20,000 to about 15 nucleotides, as all of these intermediate lengths are also contemplated to be useful, and fall within the scope of the present invention. It will be readily understood that "intermediate lengths", in these contexts, means any length between the quoted ranges, such as 14, 15, 16, 17, 18, 19, 20, *etc.*; 21, 22, 23, 24, 25, 26, 27, 28, 29, *etc.*; 30, 31, 32, 33, 34, 35, 36..... *etc.*; 40, 41, 42, 43, 44..... *etc.*, 50, 51, 52, 53..... *etc.*; 60, 61, 62, 63.... *etc.*, 70, 80, 90, 100, 110, 120, 130..... *etc.*; 200, 210, 220, 230, 240, 250..... *etc.*; including all integers in the entire range from about 14 to about 10,000, including those integers in the ranges 200-500; 500-1,000; 1,000-2,000; 2,000-3,000; 3,000-5,000 and the like.

In a preferred embodiment, the nucleic acid segments comprise a sequence of from about 1800 to about 18,000 base pair in length, and comprise one or more genes that encode an MMT polypeptide.

It will also be understood that this invention is not limited to the particular nucleic acid sequences which encode peptides of the present invention, or which encode the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, SEQ ID NO:70, SEQ ID NO:100, SEQ ID NO:102, or SEQ ID NO:108, including the DNA sequences which are particularly disclosed in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID

NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:99, SEQ ID NO:101, or SEQ ID NO:107. Recombinant vectors and isolated DNA segments may therefore variously include the peptide-coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode larger polypeptides that nevertheless include these peptide-coding regions or may encode biologically functional equivalent proteins or peptides that have variant amino acids sequences.

The DNA segments of the present invention encompass biologically functional, equivalent peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally-equivalent proteins or peptides may be created *via* the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, *e.g.*, to introduce improvements to the antigenicity of the protein or to test mutants in order to examine activity at the molecular level

If desired, one may also prepare fusion proteins and peptides, *e.g.*, where the peptide-coding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for purification or immunodetection purposes (*e.g.*, proteins that may be purified by affinity chromatography and enzyme label coding regions, respectively).

Recombinant vectors form further aspects of the present invention. Particularly useful vectors are contemplated to be those vectors in which the coding portion of the DNA segment, whether encoding a full length protein or smaller peptide, is positioned under the control of a promoter. The promoter may be in the form of the promoter that is naturally associated with a gene encoding peptides of the present invention, as may be obtained by isolating the 5′ non-coding sequences located upstream of the coding segment or exon, for example, using recombinant cloning and/or PCR™ technology, in connection with the compositions disclosed herein.

2.6 VECTORS, HOST CELLS, AND PROTEIN EXPRESSION

In other embodiments, it is contemplated that certain advantages will be gained by positioning the coding DNA segment under the control of a recombinant, or heterologous, promoter. As used herein, a recombinant or heterologous promoter is intended to refer to a promoter that is not normally associated with a DNA segment encoding an MMT polypeptide or peptide in its natural environment. Such promoters may include promoters normally associated with other genes, and/or promoters isolated from any bacterial, viral, eukaryotic, or plant cell. Naturally, it will be important to employ a promoter that effectively directs the expression of the DNA segment in the cell type, organism, or even animal, chosen for expression. The use of promoter and cell type combinations for protein expression is generally known to those of skill in the art of molecular biology; for example, see Sambrook *et al.*, 1989. The promoters employed may be constitutive, or inducible, and can be used under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins or peptides. Appropriate promoter systems contemplated for use in high-level expression include, but are not limited to, the *Pichia* expression vector system (Pharmacia LKB Biotechnology).

In connection with expression embodiments to prepare recombinant proteins and peptides, it is contemplated that longer DNA segments will most often be used, with DNA segments encoding the entire peptide sequence being most preferred. However, it will be appreciated that the use of shorter DNA segments to direct the expression of MMT peptides or epitopic core regions, such as may be used to generate anti-MMT antibodies, also falls within the scope of the invention. DNA segments that encode peptide antigens of about 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or so amino acids, and up to and including those of about 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, or 75 or so amino acids in length, or more preferably, from about 8 to about 30 amino acids in length, or even more preferably, from about 8 to about 20 amino acids in length are contemplated to be particularly useful. Such peptide epitopes may be amino acid sequences which comprise contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51.

2.7 TRANSFORMED HOST CELLS AND TRANSGENIC PLANTS

In one embodiment, the invention provides a transgenic plant having incorporated into its genome a transgene that encodes a contiguous amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51.

A further aspect of the invention is a transgenic plant having incorporated into its genome a transgene, that comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:25. Also disclosed and claimed are progeny of such a transgenic plant, as well as its seed, progeny from such seeds, and seeds arising from the second and subsequent generation plants derived from such a transgenic plant.

In yet another aspect, the present invention provides methods for producing a transgenic plant that expresses a nucleic acid segment encoding the novel recombinant MMT proteins of the present invention. The process of producing transgenic plants is well known in the art. In general, the method comprises transforming a suitable host cell with one or more DNA segments that contain one or more promoters operatively linked to a coding region that encodes one or more of the disclosed MMT proteins. Such a coding region is generally operatively linked to a transcription-terminating region, whereby the promoter is capable of driving the transcription of the coding region in the cell, and hence providing the cell the ability to produce the recombinant protein *in vivo*. Alternatively, in instances where it is desirable to control, regulate, or decrease the amount of a particular recombinant MMT expressed in a particular transgenic cell, the invention also provides for the expression of MMT-specific antisense mRNA. The use of antisense mRNA as a means of controlling or decreasing the amount of a given protein of interest in a cell is well known in the art.

Another aspect of the invention comprises a transgenic plant that expresses a gene or gene segment encoding one or more of the novel polypeptide compositions disclosed herein. As used herein, the term "transgenic plant" is intended to refer to a plant that has incorporated DNA sequences, including but not limited to genes which are perhaps not normally present, DNA sequences not normally transcribed into RNA or translated into a protein ("expressed"), or any other genes or DNA sequences which one desires to introduce into the non-transformed plant, such as genes which may normally be present in the non-transformed plant but which one desires to either genetically engineer or to have altered expression.

It is contemplated that in some instances the genome of a transgenic plant of the present invention will have been augmented through the stable introduction of one or more MMT-encoding transgenes, either native, synthetically modified, or mutated. In some instances, more than one transgene will be incorporated into the genome of the transformed host plant cell. Such is the case when more than one MMT-encoding DNA segment is incorporated into the genome of such a plant. In certain situations, it may be desirable to have one, two, three, four, or even more MMT polypeptides (either native or recombinantly-engineered) incorporated and stably expressed in the transformed transgenic plant.

A preferred gene that may be introduced includes, for example, an MMT polypeptide-encoding a DNA sequence from plant origin, such as those illustrated herein, and particularly one or more of those comprising one or more amino acid sequences described in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51.

Means for transforming a plant cell and the preparation of a transgenic cell line are well known in the art, and are discussed herein. Vectors, plasmids, cosmids, YACs (yeast artificial chromosomes) and DNA segments for use in transforming such cells will, of course, generally comprise either the operons, genes, or gene-derived sequences of the present invention, either native, or synthetically-derived, and particularly those encoding the disclosed MMT polypeptides. These DNA constructs can further include structures such as promoters, enhancers, polylinkers, or even gene sequences that have positively- or negatively-regulating activity upon the particular genes of interest as desired. The DNA segment or gene may encode

either a native or modified MMT polypeptide, which will be expressed in the resultant recombinant cells, and/or which will impart an improved phenotype to the regenerated plant, in this case, by altering or modulating the biosynthesis of sulfur-containing compounds in a transformed plant cell.

Such transgenic plants may be desirable for increasing the biosynthesis of sulfur containing compounds of a monocotyledonous or dicotyledonous plant, by incorporating into such a plant, a transgenic DNA segment encoding an MMT polypeptide. Particularly preferred plants include grains such as corn, wheat, millet, rye, rice, barley, and oats; legumes such as beans, soybeans, peas; tubers such as potatoes; fiber crops such as flax and cotton; turf and pasture grasses; tobacco, sunflower, safflower, canola, ornamental plants; shrubs; trees; vegetables, berries, citrus, fruits, cacti, succulents, and other commercially-important crops including garden, floral, and houseplants.

In a related aspect, the present invention also encompasses a seed produced by the transformed plant, a progeny from such seed, and a seed produced by the progeny of the original transgenic plant, produced in accordance with the above process. Such progeny and seeds will have one or more MMT transgene(s) stably incorporated into its genome, and such progeny plants will inherit the traits afforded by the introduction of a stable transgene in Mendelian fashion. All such transgenic plants having incorporated into their genome transgenic DNA segments encoding one or more MMT polypeptides or polypeptides are aspects of this invention. Particularly preferred transgenes for the practice of the invention include nucleic acid segments comprising one or more MMT gene(s).

2.8 Recombinant Vectors Expressing MMT-Encoding Genes

One important embodiment of the invention is a recombinant vector that comprises a nucleic acid segment encoding one or more of the novel proteins disclosed herein. Such a vector may be transferred to and replicated in a prokaryotic or eukaryotic host, with bacterial cells being particularly preferred as prokaryotic hosts, and plant cells being particularly preferred as eukaryotic hosts.

In preferred embodiments, the recombinant vector comprises a nucleic acid segment encoding the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, or a segment that encodes a polypeptide comprising one or more contiguous sequences from SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID

NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, or SEQ ID NO:51. Exemplary nucleic acid segments that encode such polypeptide sequences are those that comprise at least 31 or more contiguous nucleotides from the sequence of one or more of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO:7, as well as those nucleic acid sequences that comprise one or more of the nucleic acid sequences disclosed in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:25.

Another important embodiment of the invention is a transformed host cell that expresses one or more of these recombinant vectors. The host cell may be either prokaryotic or eukaryotic, and particularly preferred host cells are those which express the nucleic acid segment(s) comprising the recombinant vector which encode one or more MMT or MMT-derived polypeptides that comprise an at least 7 amino acid contiguous sequence from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8. Bacterial cells are particularly preferred as prokaryotic hosts, and plant cells (particularly monocot and dicot species) are particularly preferred as eukaryotic hosts

In another embodiment, the invention encompasses a method of using a nucleic acid segment of the present invention that encodes an MMT polypetide. The method generally comprises the steps of: (a) preparing a recombinant vector in which an -encoding nucleic acid segment is positioned under the control of a promoter; (b) introducing the recombinant vector into a host cell; (c) culturing the host cell under conditions effective to allow expression of the polypeptide encoded by the sequence; and (d) obtaining the expressed protein or peptide.

A wide variety of ways are available for introducing a suitable nucleic acid segment into the microorganism or eukaryotic host under conditions that allow for stable maintenance and expression of the nucleic acid segment that encodes the polypeptide. One can provide for DNA constructs that include the transcriptional and translational regulatory signals for expression of the MMT sequence, the sequence under their regulatory control and a DNA sequence homologous with a sequence in the host organism, whereby integration will occur, and/or a replication system that is functional in the host, whereby integration or stable maintenance will occur.

The transcriptional initiation signals will preferably include at least a first promoter and at least a first transcriptional initiation start site. In some instances, it may be desirable to provide for regulative expression of the enzyme, where expression of the polypeptide will only occur after release into the environment. This can be achieved with operators or a region binding to an activator or enhancers, which are capable of induction upon a change in the physical or chemical environment of the microorganisms. For example, a temperature sensitive regulatory region may be employed, where the organisms may be grown up in the laboratory without expression of a polypeptide, but upon release into the environment, expression would begin. Other techniques may employ a specific nutrient medium in the laboratory, which inhibits the expression of the polypeptide, where the nutrient medium in the environment would allow for expression of the polypeptide. For translational initiation, a ribosomal binding site and an initiation codon will be present.

Various manipulations may be employed for enhancing the expression of the messenger RNA, particularly by using an active promoter, as well as by employing sequences, which enhance the stability of the messenger RNA. The transcriptional and translational termination region will involve stop codon(s), a terminator region, and optionally, a polyadenylation signal. A hydrophobic "leader" sequence may be employed at the amino terminus of the translated polypeptide sequence in order to promote secretion of the protein across the inner membrane.

In the direction of transcription, namely in the 5′ to 3′ direction of the coding or sense sequence, the construct will involve the transcriptional regulatory region, if any, and the promoter, where the regulatory region may be either 5′ or 3′ of the promoter, the ribosomal binding site, the initiation codon, the structural gene having an open reading frame in phase with the initiation codon, the stop codon(s), the polyadenylation signal sequence, if any, and the terminator region. This sequence as a double strand may be used by itself for transformation of a microorganism host, but will usually be included with a nucleotide sequence involving a marker, where the second nucleotide sequence may be joined to the toxin expression construct during introduction of the nucleotide into the host.

By a marker is intended a structural gene which provides for selection of those hosts which have been modified or transformed. The marker will normally provide for selective advantage, for example, providing for biocide resistance, *e.g.*, resistance to antibiotics or heavy metals; complementation, so as to provide prototropy to an auxotrophic host, or the like. Preferably, complementation is employed, so that the modified host may not only be selected,

but may also be competitive in the field. One or more markers may be employed in the development of the constructs, as well as for modifying the host. The organisms may be further modified by providing for a competitive advantage against other wild-type microorganisms in the field. For example, genes expressing metal chelating agents, *e.g.*, siderophores may be introduced into the host along with the structural gene expressing the toxin. In this manner, the enhanced expression of a siderophore may provide for a competitive advantage for the -producing host, so that it may effectively compete with the wild-type microorganisms and stably occupy a niche in the environment.

Where no functional replication system is present, the construct will also include a sequence of at least 50 basepairs (bp), preferably at least about 100 bp, more preferably at least about 1000 bp, and usually not more than about 2000 bp of a sequence homologous with a sequence in the host. In this way, the probability of legitimate recombination is enhanced, so that the gene will be integrated into the host and stably maintained by the host. Desirably, the MMT gene will be in close proximity to the gene providing for complementation as well as the gene providing for the competitive advantage. Therefore, in the event that a toxin gene is lost, the resulting organism will be likely to also lose the complementing gene and/or the gene providing for the competitive advantage, so that it will be unable to compete in the environment with the gene retaining the intact construct.

A large number of transcriptional regulatory regions are available from a wide variety of microorganism hosts, such as bacteria, bacteriophage, cyanobacteria, algae, fungi, and the like. Various transcriptional regulatory regions include the regions associated with the *trp* gene, *lac* gene, *gal* gene, the λ_L and λ_R promoters, the *tac* promoter, the naturally-occurring promoters associated with the δ-endotoxin gene, where functional in the host. See for example, U. S. Patents 4,332,898; 4,342,832; and 4,356,270 (each of which is specifically incorporated herein by reference). The termination region may be the termination region normally associated with the transcriptional initiation region or a different transcriptional initiation region, so long as the two regions are compatible and functional in the host.

Where stable episomal maintenance or integration is desired, a plasmid will be employed which has a replication system that is functional in the host. The replication system may be derived from the chromosome, an episomal element normally present in the host or a different host, or a replication system from a virus that is stable in the host. A large number of plasmids are available, such as pBR322, pACYC184, RSF1010, pR01614, and the like. See for example,

Olson *et al.* (1982); Bagdasarian *et al.* (1981), Baum *et al.*, 1990, and U. S. Patents 4,356,270; 4,362,817; 4,371,625, and 5,441,884, each incorporated specifically herein by reference.

The MMT sequence can be introduced between the transcriptional and translational initiation region and the transcriptional and translational termination region, so as to be under the regulatory control of the initiation region. This construct will be included in a plasmid, which will include at least one replication system, but may include more than one, where one replication system is employed for cloning during the development of the plasmid and the second replication system is necessary for functioning in the ultimate host. In addition, one or more markers may be present, which have been described previously. Where integration is desired, the plasmid will desirably include a sequence homologous with the host genome.

The transformants can be isolated in accordance with conventional ways, usually employing a selection technique, which allows for selection of the desired organism as against unmodified organisms or transferring organisms, when present. The transformants then can be tested for enzymatic activity. If desired, unwanted or ancillary DNA sequences may be selectively removed from the recombinant bacterium by employing site-specific recombination systems, such as those described in U. S. Patent 5,441,884 (specifically incorporated herein by reference).

In accordance with the present invention, nucleic acid sequences include and are not limited to DNA, including and not limited to cDNA and genomic DNA, genes; RNA, including and not limited to mRNA and tRNA; antisense sequences, PNAs (peptide nucleic acids), nucleosides, and suitable nucleic acid sequences such as those set forth herein, as well as alterations in the nucleic acid sequences including alterations, deletions, mutations, and homologs capable of expressing the MMT polypeptides and peptide fragments of the present invention.

As such the present invention also concerns DNA segments, that are free from total genomic DNA and that encode the novel plant-derived polypeptides disclosed herein. DNA segments encoding these peptide species may prove to encode proteins, polypeptides, subunits, functional domains, and the like of MMT-related or other non-related gene products. In addition these DNA segments may be synthesized entirely *in vitro* using methods that are well known to those of skill in the art.

As used herein, the term "DNA segment" refers to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a MMT

polypeptide or peptide refers to a DNA segment that contains MMT coding sequences yet is isolated away from, or purified free from, total genomic DNA of the species from which the DNA segment was obtained. Included within the term "DNA segment", are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phagemids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified MMT polypeptide-encoding gene refers to a DNA segment which may include in addition to peptide encoding sequences, certain other elements such as, regulatory sequences, isolated substantially away from other naturally occurring genes or protein-encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein-, polypeptide- or peptide-encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, operon sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides or peptides.

"Isolated substantially away from other coding sequences" means that the gene of interest, in this case, a gene encoding a plant MMT polypeptide, forms the significant part of the coding region of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or operon coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes, recombinant genes, synthetic linkers, or coding regions later added to the segment by the hand of man.

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5′ or 3′ sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein activity where MMT protein expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5′ or 3′ portions of the MMT coding region or may include various internal sequences, *i.e.*, introns, which are known to occur within eukaryotic genes.

The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore

contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

Highly preferred nucleic acid segments of the present invention comprise one or more MMT-encoding polynucleotide sequences of the invention, or a portion of one or more such sequences. For certain application, relatively small contiguous nucleic acid sequences are preferable, such as those which are about 14 or 15 or 16 or 17 or 18 or 19, or 20, or 30-50, 51-80, 81-100 or so nucleotides in length. Alternatively, in some embodiments, and particularly those involving preparation of recombinant vectors, transformation of suitable host cells, and preparation of transgenic plant cell, longer nucleic acid segments are preferred, particularly those that include the entire coding region of one or more MMT-encoding nucleic acid segments. As such, the preferred segments may include those that are up to about 20,000 or so nucleotides in length, or alternatively, shorter sequences such as those about 19,000, about 18,000, about 17,000, about 16,000, about 15,000, about 14,000, about 13,000, about 12,000, 11,000, about 10,000, about 9,000, about 8,000, about 7,000, about 6,000, about 5,000, about 4,500, about 4,000, about 3,500, about 3,000, about 2,500, about 2,000, about 1,500, about 1,000, about 500, or about 200 or so base pairs in length. Of course, these numbers are not intended to be exclusionary of all possible intermediate lengths in the range of from about 20,000 to about 15 nucleotides, as all of these intermediate lengths are also contemplated to be useful, and fall within the scope of the present invention. It will be readily understood that "intermediate lengths", in these contexts, means any length between the quoted ranges, such as 14, 15, 16, 17, 18, 19, 20, *etc.*; 21, 22, 23, 24, 25, 26, 27, 28, 29, *etc.*; 30, 31, 32, 33, 34, 35, 36.....*etc.*; 40, 41, 42, 43, 44.....*etc.*, 50, 51, 52, 53.....*etc.*; 60, 61, 62, 63.... *etc.*, 70, 80, 90, 100, 110, 120, 130.....*etc.*; 200, 210, 220, 230, 240, 250.....*etc.*; including all integers in the entire range from about 14 to about 10,000, including those integers in the ranges 200-500; 500-1,000; 1,000-2,000; 2,000-3,000; 3,000-5,000 and the like.

In a preferred embodiment, the nucleic acid segments comprise a sequence of from about 31 to about 3500 base pairs or so in length, and comprise at least a first sequence region that encodes all, or substantially all of a plant-derived MMT polypeptide.

Recombinant vectors and isolated DNA segments may therefore variously include the peptide-coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode larger polypeptides that nevertheless include

these peptide-coding regions or may encode biologically functional equivalent proteins or peptides that have variant amino acids sequences.

The DNA segments of the present invention encompass biologically-functional, equivalent peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally-equivalent proteins or peptides may be created *via* the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, *e.g.*, to introduce improvements to the antigenicity of the protein or to test mutants in order to examine activity at the molecular level

If desired, one may also prepare fusion proteins and peptides, *e.g.*, where the peptide-coding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for purification or immunodetection purposes (*e.g.*, proteins that may be purified by affinity chromatography and enzyme label coding regions, respectively).

Recombinant vectors form further aspects of the present invention. Particularly useful vectors are contemplated to be those vectors in which the coding portion of the DNA segment, whether encoding a full-length protein or smaller peptide, is positioned under the control of a promoter. The promoter may be in the form of the promoter that is naturally associated with a gene encoding peptides of the present invention, as may be obtained by isolating the 5′ non-coding sequences located upstream of the coding segment or exon, for example, using recombinant cloning and/or PCR™ technology, in connection with the compositions disclosed herein.

2.9 Transformed Host Cells and Transgenic Plants

In one embodiment, the invention provides a transgenic plant having incorporated into its genome a transgene that encodes at least a 7-amino acid contiguous sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.

A further aspect of the invention is a transgenic plant having incorporated into its genome a selected nucleic acid sequence that comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO:7. Also disclosed and claimed are the progeny or offspring of such a transgenic plant, as well as its fruit, nuts,

and/or seed, progeny from such fruit, nuts, and/or seeds, as well as all fruits, nuts and/or seeds arising from the second and all subsequent generation plants derived from such a parental transgenic plant, plant tissue or transformed plant host cell. The invention also discloses and claims host cells, both native, and genetically engineered, which express the novel MMT-encoding sequence to produce polypeptides having MMT enzymatic activity.

Methods of using such cells to produce an MMT polypeptide are also disclosed. Such methods generally involve culturing the host cell under conditions effective to produce an MMT polypeptide, and obtaining the polypeptide so produced from said cell.

In yet another aspect, the present invention provides methods for producing a transgenic plant that expresses a nucleic acid segment encoding the novel MMT polypeptides of the present invention. The process of producing transgenic plants is well known in the art. In general, the method comprises transforming a suitable host cell with one or more nucleic acid segments that contain one or more promoters operatively linked to a coding region that encodes one or more of the disclosed MMT proteins. Such a coding region is generally operatively linked to a transcription-terminating region, whereby the promoter is capable of driving the transcription of the coding region in the cell, and hence providing the cell the ability to produce the recombinant protein *in vivo*. Alternatively, in instances where it is desirable to control, regulate, or decrease the amount of a particular recombinant MMT polypeptide expressed in a particular transgenic cell, the invention also provides for the expression of MMT antisense mRNA. The use of antisense mRNA as a means of controlling or decreasing the amount of a given protein of interest in a cell is well known in the art.

Another aspect of the invention comprises a transgenic plant that expresses a gene or gene segment encoding one or more of the novel MMT polypeptide compositions disclosed herein. As used herein, the term "transgenic plant" is intended to refer to a plant that has stably incorporated DNA sequences, including but not limited to genes that are perhaps not normally present, DNA sequences not normally transcribed into RNA or translated into a protein ("expressed"), or any other genes or DNA sequences that one desires to introduce into the non-transformed plant, such as genes which may normally be present in the non-transformed plant but which one desires to either genetically engineer or to have altered expression.

It is contemplated that in some instances the genome of a transgenic plant of the present invention will have been augmented through the stable introduction of one or more MMT-encoding transgenes, either native, synthetically modified, or mutated. In some instances, more

than one transgene will be incorporated into the genome of the transformed host plant cell. Such is the case when more than one MMT-encoding DNA segment is incorporated into the genome of such a plant. In certain situations, it may be desirable to have one, two, three, four, or even more MMT proteins (either native or recombinantly-engineered) incorporated and stably expressed in the transformed transgenic plant.

A preferred selected nucleic acid sequence that may be introduced into a target host plant includes, for example, a polynucleotide that encodes an MMT polypeptide, and particularly one or more of those described in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8. Highly preferred nucleic acid sequences are those obtained from MMT-expressing plants, or any of those sequences that have been genetically engineered to decrease or increase the enzymatic activity of the MMT protein in such a transformed host cell.

Means for transforming a plant cell and the preparation of a transgenic cell line are well known in the art, and are discussed herein. Vectors, plasmids, cosmids, YACs (yeast artificial chromosomes) and DNA segments for use in transforming such cells will, of course, generally comprise either the operons, genes, or gene-derived sequences of the present invention, either native, or synthetically-derived, and particularly those encoding the disclosed MMT polypeptides and proteins. These DNA constructs can further include structures such as promoters, enhancers, polylinkers, or even gene sequences that have positively- or negatively-regulating activity upon the particular genes of interest as desired. The DNA segment or gene may encode either a native or modified MMT polypeptide or protein, which will be expressed in the resultant recombinant cells, and/or which will impart an improved phenotype to the regenerated plant

Such transgenic plants may be desirable for increasing the biosynthesis of sulfur-containing compounds in a monocotyledonous or dicotyledonous plant, by incorporating into such a plant, a transgenic DNA segment encoding at least a first MMT polypeptide. Particularly preferred plants include grains such as safflower, sunflower, tobacco, corn, wheat, rye, millet, rice, barley, and oats; legumes such as beans, peas, soybeans; tubers such as potatoes; fiber crops such as flax and cotton; turf and pasture grasses; ornamental plants; shrubs; trees; vegetables, berries, citrus, fruits, cacti, succulents, and other commercially-important crops including garden and houseplants.

In a related aspect, the present invention also encompasses a seed produced by the transformed plant, a progeny from such seed, and a seed produced by the progeny of the original transgenic plant, produced in accordance with the above process. Such progeny and seeds will

have one or more MMT-encoding transgene(s) stably incorporated into its genome, and such progeny plants will inherit the traits afforded by the introduction of a stable transgene in Mendelian fashion. All such transgenic plants having incorporated into their genome transgenic DNA segments encoding one or more MMT polypeptides are aspects of this invention. Particularly preferred transgenes for the practice of the invention include nucleic acid segments comprising one or more nucleic acid sequences that encode an MMT polypeptide.

3.0 BRIEF DESCRIPTION OF THE DRAWINGS

The drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1 shows the SMM cycle of flowering plants.

FIG. 2A and FIG 2B provide evidence that SMM occurs in phloem exudates from diverse plants. FIG. 2A shows the [^{35}S]SMM contents of exudates, as a percentage of total ^{35}S exuded. Attached leaves were supplied with 20 μCi (17 pmol) of L-[^{35}S]Met, severed after 2 h, and placed with their cut ends in 5 mM Na_2EDTA, pH 7.0, for 20 h in darkness. Total ^{35}S exudation ranged from 7.6 nCi (maize) to 917 nCi (*Arabidopsis*). Inset is an autoradiograph of a TLC separation of the BioRex-70 fraction of exudates from representative species: wheat (W); canola (C); and soybean (S). FIG. 2B shows the levels of amino acids exuded with or without 5 mM Na_2EDTA in the medium. Leaves were matched in size and age to those supplied with [^{35}S]Met. Values for zucchini have been multiplied by 0.5 to fit the scale used for other species.

FIG. 3A, FIG. 3B and FIG. 3C provide the deduced amino acid sequences of *W. biflora*, *Arabidopsis*, and maize MMTs. The *W. biflora* sequence is a composite of the longest (3.1 kb) cDNA and a 5' RACE product; these overlapped by 517 bp and were identical in the overlap region. Peptide sequences from purified *W. biflora* MMT are underlined. Identical residues are in bold, similar residues are boxed. Wb, *W. biflora;* Zm, maize; and At, *Arabidopsis*.

FIG. 4 illustrates a possible role of the SMM cycle in long-distance sulfur transport. The scheme shows a major flux (wide arrows) from Met to SMM in leaves, phloem transport of SMM, and reconversion of SMM to Met in developing seeds or other sinks.

FIG. 5A, FIG. 5B, FIG. 5C, FIG. 5D, FIG. 5E and FIG. 5F provide a comparison of SEQ ID NO:1 *vs.* SEQ ID NO:5 indicating particular homologous regions shared between the two sequences that have ≥12 contiguous nucleotides of 100% identical sequence. The areas of identity were determined using the BLAST v2.0.9 computer algorithm available from the National Center for Biotechnology Information.

FIG. 6A, FIG. 6B and FIG. 6C provide a comparison of SEQ ID NO:2 *vs.* SEQ ID NO:6 indicating particular homologous regions shared between the two sequences that have ≥8 contiguous amino acids of 100% identity. The areas of identity were determined as in FIG. 5A.

FIG. 7A and FIG. 7B provide a comparison of SEQ ID NO:7 *vs.* SEQ ID NO:5 indicating particular homologous regions shared between the two sequences that have ≥14 contiguous nucleotides of 100% identity. The areas of identity were determined as in FIG. 5A.

FIG. 8A and FIG. 8B provide a comparison of SEQ ID NO:8 *vs.* SEQ ID NO:6 indicating particular homologous regions shared between the two sequences that have ≥11 contiguous amino acids of 100% identity. The areas of identity were determined as in FIG. 5A.

4.0 DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Compositions and methods for modulating the biosynthesis of organic sulfur compounds in plants, particularly sulfur amino acids, more particularly cysteine and methionine are provided. The methods involve transforming a plant with at least one nucleotide sequence of the invention.

In particular, the present invention provides for isolated nucleic acid molecules comprising nucleotide sequences encoding the amino acid sequences shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8 or the nucleotide sequences encoding these polypeptides, and particularly those shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO:7. Antisense sequences for the MMT sequences of the invention are also provided. Further provided are polypeptides having an amino acid sequence encoded by a nucleic acid molecule described herein, for example those set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 and SEQ ID NO:7.

Fragments and variants of the disclosed nucleotide sequences and proteins encoded thereby are also encompassed by the present invention. By "fragment" is intended a portion of the nucleotide sequence or a portion of the amino acid sequence and hence protein encoded thereby. Fragments of a nucleotide sequence may encode protein fragments that retain the

biological activity of the native protein. Alternatively, fragments of a nucleotide sequence that are useful as hybridization probes generally do not encode fragment proteins retaining biological activity. Thus, fragments of a nucleotide sequence may range from at least about 20 nucleotides, about 50 nucleotides, about 100 nucleotides, and up to the entire nucleotide sequence encoding the proteins of the invention.

A fragment of an MMT nucleotide sequence that encodes a biologically active portion of an MMT protein of the invention will encode at least 15, 25, 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, or 1000 contiguous amino acids, or up to the total number of amino acids present in a full-length MMT protein of the invention (for example, 1088, 1071, or 1091 amino acids for SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, respectively). Fragments of the MMT nucleotide sequence useful as hybridization probes for PCR™ primers generally need not encode a biologically active portion of an MMT protein.

A fragment of an MMT nucleotide sequence may encode a biologically active portion of an MMT protein, or it may be a fragment that can be used as a hybridization probe or PCR™ primer using methods disclosed below. A biologically active portion of an MMT protein can be prepared by isolating a portion of one of the MMT nucleotide sequences of the invention, expressing the encoded portion of the MMT protein (e.g., by recombinant expression *in vitro*), and assessing the activity of the encoded portion of the MMT protein. Nucleic acid molecules that are fragments of an MMT nucleotide sequence comprise at least about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, and about 30 or so contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7. Slightly longer sequences include those that comprise at least about 31, about 32, about 33, about 34, about 35, about 36, about 37, about 38, about 39, about 40, about 41, about 42, about 43, about 44, about 45, about 46, about 47, about 48, about 49, or about 50 or so contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7. Still longer sequences include those that comprise at least about 51, about 52, about 53, about 54, about 55, about 56, about 57, about 58, about 59, about 60, about 61, about 62, about 63, about 64, about 65, about 6, about 67, about 68, about 69, or about 70 or so contiguous nucleotides from any one of or so contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7. When it is desirable to identify even longer segments that comprise still longer contiguous nucleic acid sequences from or so contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or

SEQ ID NO:7, one may prepare polynucleotides that comprise about 75, about 80, about 85, about 90, about 95, about 100, about 125, about 150, about 175, about 200, about 225, about 250, about 275, about 300, about 325, about 350, about 375, about 400, about 425, about 450, abtou 500, about 550, about 600, about 650, about 700, about 750, about 800, about 850, about 900, about 950, about 1000, about 1050, about 1100, about 1150, about 1200, about 1250, about 1300, about 1350, about 1400, about 1450, about 1500, about 1550, about 1600, about 1650, about 1700, about 1750, about 2000, about 2250, about 2500, about 2750, about 3000, about 3250, or about 3500 nucleotides, or so nucleotides, and even those comprising up to and including the number of nucleotides present in a full-length MMT nucleotide sequence disclosed herein (for example, 3646, 3558, or 3553 nucleotides for the sequences illustrated in SEQ ID NO:1, SEQ ID NO:3, and SEQ ID NO:5, respectively).

By "variants" are intended substantially similar sequences. For nucleotide sequences, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the amino acid sequence of an MMT protein. Generally, nucleotide sequence variants of the invention will have at least 40%, 50%, 60%, 70%, generally, 80%, preferably 90%, 95%, 98% sequence identity to its respective native nucleotide sequence.

By "variant" protein is intended a protein derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

These nucleotide sequences can be used to isolate other homologous sequences, particularly from other plant species. Methods are readily available in the art for the hybridization of nucleic acid sequences. Coding sequences from other plants may be isolated according to well-known techniques based on their sequence homology to the coding sequences set forth herein. In these techniques all or part of the known coding sequence is used as a probe that selectively hybridizes to other MMT coding sequences present in a population of cloned genomic DNA fragments or cDNA fragments (*i.e.* genomic or cDNA libraries) from a chosen organism.

To obtain other sequences, the entire MMT sequence or portions thereof may be used as probes capable of specifically hybridizing to corresponding coding sequences and messenger

RNAs. To achieve specific hybridization under a variety of conditions, such probes include sequences that are unique and are preferably at least about 10 nucleotides in length, and most preferably at least about 20 nucleotides in length. Such probes may be used to amplify the MMT coding sequences of interest from a chosen organism by the well-known process of polymerase chain reaction (PCR™). This technique may be used to isolate additional coding sequences from a desired organism or as a diagnostic assay to determine the presence of coding sequences in an organism.

Such techniques include hybridization screening of plated DNA libraries (either plaques or colonies) (Sambrook *et al.*, 1989) and amplification by PCR™ using oligonucleotide primers corresponding to sequence domains conserved among the amino acid sequences (Innis *et al.*, 1990).

For example, hybridization of such sequences may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions (*e.g.*, conditions represented by a wash stringency of 35-40% Formamide with 5X Denhardt's solution, 0.5% SDS and 1X SSPE at 37°C; conditions represented by a wash stringency of 40-45% Formamide with 5X Denhardt's solution, 0.5% SDS, and 1X SSPE at 42°C; and conditions represented by a wash stringency of 50% Formamide with 5X Denhardt's solution, 0.5% SDS and 1X SSPE at 42°C, respectively), to DNA encoding the MMT proteins disclosed herein in a standard hybridization assay (Sambrook *et al.*, 1989). In general, polynucleotide sequences which encode an MMT polypeptide as disclosed herein and which hybridize to one or more of the polynucleotide sequences disclosed herein will be at least 50% homologous, 70% homologous, and even 85% homologous or more with the disclosed sequence. That is, the sequence similarity of sequences may range, sharing at least about 50%, about 70%, and even about 85% sequence similarity.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", (d) "percentage of sequence identity", and (e) "substantial identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a, polynucleotide sequence, wherein the polynucteotide sequence in the comparison

window may comprise additions or deletions (*i.e.* gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith *et al.*, (1981); by the homology alignment algorithm of Needleman *et al.*, (1970); by the search for similarity method of Pearson *et al.*, (1988); by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California; GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA; the CLUSTAL program is well described by Higgins *et al.*, (1988); Higgins *et al.*, (1989); Corpet *et al.*, (1988); Huang *et al.*, (1992), and Person *et al.*, (1994); preferred computer alignment methods also include the BLASTP, BLASTN, and BLASTX algorithms (Altschul *et al.*, 1990). Alignment is also often performed by inspection and manual alignment. Using the Align Plus program and parameters set as mismatch of 2, open gap of 4, and extend gap of 1, the disclosed nucleotide sequences were found to be 63% and 67% homologous.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (*e.g.*, charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an

identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, CA).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.* gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70% sequence identity, preferably at least 80%, more preferably at least 90%, and most preferably at least 95%, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 60%, more preferably at least 70%, 80%, 90%, and most preferably at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Generally, stringent conditions are selected to be about 5°C to about 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent wash conditions are those in which the salt concentration is about 0.02 M at pH 7 and the temperature is at least about 50°C, about 55°C, or even about 60°C or so. However, nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy

permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70% sequence identity to a reference sequence, preferably 80%, more preferably 85%, most preferably at least 90% or 95% sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman *et al.*, (1970). An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. Peptides that are "substantially similar" share sequences as noted above except that residue positions that are not identical may differ by conservative amino acid changes.

The proteins of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertion. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the proteins can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art (Kunkel, 1985; Kunkel *et al.*, 1987; U.S. Patent No. 4,873,192; Walker and Gaastra, 1983).

Thus, the genes and nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant forms. Likewise, the proteins of the invention encompass both naturally occurring proteins as well as variations and modified forms thereof. Such variants will continue to possess the desired MMT activity. Obviously, the mutations that will be made in the DNA encoding the variant must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure (see *e.g.*, European Patent Application Publication No. 75,444, specifically incorporated herein by reference in its entirety).

"Organic sulfur compounds" are intended to include, but not be limited to, biochemical compounds such as glutathione, phytochelatins, sulfur-containing vitamins, glucosinotates, dimethylsulfoniopropionate, amino acids such as methionine and cysteine, and the like.

MMT activity has been found in flowering plants (Giovanelli *et al.*, 1980; Mudd *et al.*, 1990). It has been purified from, and quantitated in, leaves of *Wollastonia biflora* (James *et al.*,

1995a; 1995b) and from germinating barley (Pimenta *et al.*, 1998) using the methods and assays described in those references. The estimated size of MMT is about three times larger than any other small-molecule methyltransferase (Fujioka, 1992; Joshi *et al.*, 1998).

Although it is clear how plants synthesize SMM and reconvert it to Met, the physiological role(s) of SMM and its cycle have remained largely undefined until the present invention. Proposed general roles for SMM and its cycle include Met storage, methyl donation, and regulation of the AdoMet/Met ratio, all of which are reasonable but as yet unsupported by experimental evidence (Giovanelli *et al.*, 1980; Mudd *et al.*, 1990). In all these proposed roles, SMM would exert its functions without exiting the cells that produce it.

While the invention is not bound by any particular mechanism of action, it is proposed that SMM has a role in phloem sulfur transport. SMM is a major sulfur- long distance transport compound in plants. Furthermore, SMM is ubiquitous in higher plants. Therefore, by the methods of the invention, increasing the biosynthesis of SMM in the plant leads to increased levels of SMM in sinks, *e.g.*, seeds.

In this manner, the methods of the invention comprise transforming a plant with an MMT sequence to modulate organic sulfur metabolism. By "modulate" is intended an increase or decrease in activity. Generally, to increase biosynthesis, the plant is transformed with a DNA construct comprising an MMT coding sequence. To decrease biosynthesis, an antisense construct can be used.

Other nucleic acids encoding enzymes involved in organic sulfur compound biosynthesis, such as cysteine and methionine biosynthesis can be utilized to shunt the pathway in particular directions. For example, in seeds, SMM would be converted to methionine by HMT. By increasing the expression of HMT in the seeds, the pathway could be shunted into methionine in seeds.

In the same manner, antisense constructs for enzymes involved in the biosynthesis of sulfur compounds could be utilized to direct biosynthesis into a particular product or to stop biosynthesis for the build-up of a particular compound. Likewise, the compositions and methods of the invention can be used in combination with other methods to modulate sulfur compounds in plants.

Enzymes involved in cysteine and methionine biosynthesis are known in the art. See, *e.g.*, aspartokinase (Masakazu *et al.*, 1992; Japanese Patent number 1994062866-A 1, GenBank Accession No. E06825; Omori *et al.*, 1993; GenBank Accession No. X60821; Moriya *et al.*,

1995; Japanese Patent number 1997070291-A; GenBank Accession No. E 12770); aspartate semialdehyde dehydrogenase (Calzada, F.R.A., direct submission, GenBank Accession No. Y15281; Daniel *et al.*, 1993; GenBank Accession No. Z22554; Chen *et al.*, 1993; GenBank Accession No. Z22554; GenBank Accession No. U90239; Brakhage *et al.*, 1990; GenBank Accession No. Z75208; Gothel *et al.*, 1997; GenBank Accession No. Z75208); homoserine kinase (see also aspartokinase, GenBank Accession No. X60821; Nakabachi *et al.*, 1997; GenBank Accession No. AB004856; Ryoichi *et al.*, 1986; Japanese Patent number 1987232392-A 1 and Japanese Patent number JP 1986076298; GenBank Accession No. E01358; Sadao *et al.*, Japanese Patent number 1993207886-A 4; GenBank Accession No. D14072); threonine synthase (see also aspartokinase, GenBank Accession No. X6082; GenBank Accession No. Z46263; Rognes, S.E., direct submission, GenBank Accession No. Z46263; GenBank Accession No. L41666; Clepet *et al.*, 1992; GenBank Accession Nos. X65033 and S50569; Cami, B., direct submission, GenBank Accession Nos. X65033 and S50569); cystathionine gamma synthase (Kim and Leustek, 1996; GenBank Accession No. AF069317; Locke *et al.*, direct submission, GenBank Accession No. AF007786); cystathionine beta lyase (Bork *et al.*, 1997; GenBank Accession No. AJ001148; Sienko, M., direct submission, GenBank Accession No. U28383; Ravanel *et al.*, 1995; GenBank Accession No. L40511); methionine synthase (Kurvari *et al.*, 1995; GenBank Accession No. U36197; Ravanel *et al.*, 1998; GenBank Accession No. U97200; Michalowski *et al.*, direct submission, GenBank Accession No. U84889; Eichel *et al.*, 1995; GenBank Accession No. X83499); ATP sulfurylase (Murillo *et al.*, 1995; GenBank Accession No. U06275; Leustek *et al.*, 1994; GenBank Accession No. U05218; Bolchi *et al.*, direct submission, GenBank Accession No. AF016305; Laue *et al.*, 1994; GenBank Accession No. L26897; Laeremans *et al.*, GenBank Accession No. AJ001223); APS kinase (Korch *et al.*, 1991; GenBank Accession No. S55315; Arz *et al.*, 1994; GenBank Accession No. AF044285; Schiffmann *et al.*,1998; GenBank Accession No. AF044285, GenBank Accession No. AF043351, PGR98-116, GenBank Accession No. AF044285; Jain *et al.*, 1994; GenBank Accession No. U05238; Lee *et al.*, 1998; GenBank Accession No. U05238); APS reductase (Speich *et al.*, 1994; GenBank Accession No. Z69372; Setya *et al.*, 1996; GenBank Accession No. U56921; Bick *et al.*, 1998; GenBank Accession No. U56921); PAPS reductase (Krone *et al.*, 1991; GenBank Accession No. Y07525; Krone *et al.*, 1990; GenBank Accession No. Y07525; Gutierrez-Marcos *et al.*, 1996; GenBank Accession No. U53865; Schwenn, J.D., direct submission, GenBank Accession No. Z23169); see number five under ATP sulfurylase,

GenBank Accession No. AJ001223; Bussey *et al.*, 1997; GenBank Accession Nos. U25840 and U00094); sulfate reductase (GenBank Accession No. Y07525; GenBank Accession No. Z23169; Hipp *et al.*, 1997; GenBank Accession No. U84760; Pott *et al.*, 1998; GenBank Accession No. U84760; Bork *et al.*, 1998; GenBank Accession No. Y10157; Mbeguie-A-Mbeguie *et al.*, GenBank Accession No. AF071890; Bruehl *et al.*, 1996; GenBank Accession No. Z49217; Hummerjohann *et al.*, 1998; GenBank Accession No. AF026066); serine acetyltransferase (GenBank Accession No. X80938; GenBank Accession No. D88529; Saito *et al.*, 1995; GenBank Accession No. D49535); cysteine synthase (Hesse *et al.*, 1998; GenBank Accession No. AF044172) and plastidic (GenBank Accession No. AF044173) cysteine synthase isoforms from *Solanum tuberosum* (PGR98-057); GenBank Accession No. AF044173; Brander *et al.*, 1995; GenBank Accession No. X85803; Topczewski *et al.*, 1997; GenBank Accession No. U19395); gamma glutamylcysteine synthase (Powles *et al.*, 1996; GenBank Accession No. U81808 L75931; GenBank Accession No. AL031018; EU, *Arabidopsis* sequencing project, direct submission, GenBank Accession No. AL031018); glutathione synthetase (Okumura *et al.*, 1997; GenBank Accession No. D88540; Inoue *et al.*, 1998; GenBank Accession No. Y13804; Accession No. Y10984; and GenBank Accession No. U22359).

Variants and functional fragments, including shufflents, of the above enzymes or of MMT or of other enzymes of the SMM cycle, may also be utilized. It is only required that the enzymes have an activity sufficient to modulate the level of a particular organic sulfur compound in a plant. Variants can be produced by methods known in the art. Variant proteins include those proteins derived from the native protein by deletion (so-called truncation), addition, or substitution of one or more amino acids at one or more sites in the native protein.

For example, amino acid sequence variants of the polypeptide can be prepared by mutations in the cloned DNA sequence encoding the native protein of interest. Methods for mutagenesis and nucleotide sequence alterations are well known in the art (Walker-and Gaastra, 1983; Kunkel, 1985; Kunkel *et al.*, 1987; Sambrook *et al.*, 1989; U.S. Patent No. 4,873,192; and the references cited therein; specifically incorporated herein by reference in its entirety. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff *et al.*, (1978), herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be preferred.

The MMT nucleic acids, as well as any additional genes of interest, can be optimized for enhanced expression in plants of interest (Eur. Pat. Appl. Ser. No. EP A0359472; Intl. Pat. Appl. Publ. Ser. No. W0 91/16432; Perlak *et al.*, 1991; Murray *et al.*, 1989). In this manner, the nucleic acids can be synthesized utilizing plant-preferred codons, (Murray *et al.*, 1989), the disclosure of which is incorporated herein by reference. In this manner, synthetic nucleic acids can also be made based on the distribution of codons a particular host uses for a particular amino acid.

Another method for obtaining modified enzymes that can alter the level of at least one organic sulfur compound is by sequence shuffling. Sequence shuffling has been described, *e.g.*, in Intl. Pat. Appl. Publ. Ser. No. WO 96/19256; Zhang *et al.*, 1997). Libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*.

In some instances, the enzymes of interest are natively expressed in the plant. However, by transformation with heterologous promoters, expression levels or patterns can be altered (Sambrook *et al.*, 1989; Innis *et al.*, 1990).

The nucleic acids can be combined with constitutive, tissue-specific, or other promoters for expression in plants. Such constitutive promoters include, for example, the core CaMV 35S promoter (Odell *et al.*, 1985); rice actin (McElroy *et al.*, 1990); ubiquitin (Christensen *et al.*, 1989; Christensen *et al.*, 1992); pEMU (Last *et al.*, 1991); MAS (Velten *et al.*, 1984), and the like. Other constitutive promoters have been described, *e.g.*, in U.S. Patent No. 5,608,149; U.S. Patent No. 5,608,144; U.S. Patent No. 5,604,121; U.S. Patent No. 5,569,597; U.S. Patent No. 5,466,785; U.S. Patent No. 5,399,680; U.S. Patent No. 5,268,463; and U.S. Patent No. 5,608,142, each of which is specifically incorporated herein by reference in its entirety.

"Seed-specific" promoters of the invention include embryo-specific promoters. Additionally, such promoters include globulin 1, cruciferin, napin, B-conglycinin, phaseolin, as well as other promoters associated with storage proteins or involved in fatty acid biosynthesis.

The polynucleotides of the invention may be provided in expression cassettes to facilitate expression in plants. Expression cassettes will comprise a transcriptional initiation region linked to the coding sequence or antisense sequence of the nucleotide of interest. Such an expression cassette is generally provided with a plurality of restriction sites for insertion of the sequence to

be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

The transcriptional initiation region, the promoter, may be native (*i.e.* analogous) or foreign (*i.e.* heterologous) to the plant host. Additionally, the promoter may be a synthetic sequence. By foreign is intended that the transcriptional initiation region is not found in the native plant into which the transcriptional initiation region is introduced.

The transcriptional cassette will include in the 5' to 3' direction of transcription, a transcriptional and translational initiation region, a DNA sequence of interest, and a transcriptional and translational termination region functional in plants. The termination region may be native with the transcriptional initiation region, may be native with the DNA sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens* such as the octopine synthase and nopaline synthase termination regions (Guerineau *et al.*, 1991; Proudfoot, 1991; Sanfacon *et al.*, 1991; Mogen *et al.*, 1990; Munroe *et al.*, 1990; Ballas *et al.*, 1989; Joshi *et al.*, 1987).

In preparing the expression cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resubstitutions, *e.g.*, transitions and transversions, may be involved.

The biosynthesis of organic sulfur compounds can be altered in any plant of interest. Of particular interest are plants useful for human and domestic animal food. Such plants include forages and seed crop plants such as cereal crops and oil seed crops. Of particular interest are plants where the seed is produced in high amounts, or the seed or a seed part is edible. Seeds of interest include the oil seeds, such as seeds from *Brassica*, cotton, soybean, safflower, sunflower, coconut, palm, *etc.;* grain seeds such as wheat, rice, com, *etc.;* other seeds including oats, pumpkin, squash, poppy, sesame, peanut, peas, beans, cocoa, coffee, *etc.*; and tree nuts such as walnuts, pecans, almonds, *etc*. Especially preferred plants are corn, soybean, sunflower, *Brassica*, wheat, rye, rice, millet, sorghum, and alfalfa.

The modified plant may be grown into plants in accordance with conventional ways (McCormick *et al.*, 1986). These plants may then be grown, and either pollinated with the same

transformed strain or different strains, and the resulting hybrid having the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that the subject phenotypic characteristic is stably maintained and inherited into the progeny and subsequent generations of the transformed plant. Likewise, the seeds from the transformed plant or from a progeny or subsequent generation of the plant may be harvested and assayed to ensure the desired phenotype has been achieved in the progeny and the seeds from the transgenic plant and its offspring.

4.1 Transformed Host Cells and Transgenic Plants

A bacterial cell, a yeast cell, or a plant cell transformed with an MMT-encoding gene-containing expression vector of the present invention also represents an important aspect of the present invention. Furthermore, transgenic plants and the progeny and seeds derived from such a transformed or transgenic plant are also important aspects of this invention.

Such transformed host cells are often desirable for use in the expression of the various DNA gene constructs disclosed herein. In some aspects of the invention, it is often desirable to modulate, regulate, or otherwise control the expression of the gene segments disclosed herein. Such methods are routine to those of skill in the molecular genetic arts. Typically, when increased or over-expression of a particular gene is desired, various manipulations may be employed for enhancing the expression of the messenger RNA, particularly by using an active promoter, as well as by employing sequences, which enhance the stability of the messenger RNA in the particular transformed host cell.

Typically, the initiation and translational termination region will involve stop codon(s), a terminator region, and optionally, a polyadenylation signal. In the direction of transcription, namely in the 5′ to 3′ direction of the coding or sense sequence, the construct will involve the transcriptional regulatory region, if any, and the promoter, where the regulatory region may be either 5′ or 3′ of the promoter, the ribosomal binding site, the initiation codon, the structural gene having an open reading frame in phase with the initiation codon, the stop codon(s), the polyadenylation signal sequence, if any, and the terminator region. This sequence as a double strand may be used by itself for transformation of a microorganism host, but will usually be included with a DNA sequence involving a marker, where the second DNA sequence may be joined to the expression construct during introduction of the DNA into the host.

Where no functional replication system is present, the construct will also include a sequence of at least 50 basepairs (bp), preferably at least about 100 bp, and usually not more than about 1000 bp of a sequence homologous with a sequence in the host. In this way, the probability of legitimate recombination is enhanced, so that the gene will be integrated into the host and stably maintained by the host. Desirably, the MMT-encoding gene-promoter construct will be in close proximity to the gene providing for complementation as well as the gene providing for the competitive advantage. Therefore, in the event that an MMT-encoding gene is lost, the resulting organism will be likely to also lose the complementing gene and/or the gene providing for the competitive advantage, so that it will be unable to compete in the environment with the gene retaining the intact construct.

The MMT-encoding gene can be introduced between the transcriptional and translational initiation region and the transcriptional and translational termination region, so as to be under the regulatory control of the initiation region. This construct will be included in a plasmid, which will include at least one replication system, but may include more than one, where one replication system is employed for cloning during the development of the plasmid and the second replication system is necessary for functioning in the ultimate host. In addition, one or more markers may be present, which have been described previously. Where integration is desired, the plasmid will desirably include a sequence homologous with the host genome.

Alternatively, the left and right T-DNA borders from the Ti plasmid may be used when integration is desired using *A. tumefaciens* vectors for plant transformation. The transformants can be isolated in accordance with conventional ways, usually employing a selection technique, which allows for selection of the desired organism as against unmodified organisms or transferring organisms, when present. The transformants then can be tested for presence of the genetic construct.

Genes or other nucleic acid segments, as disclosed herein, can be inserted into host cells using a variety of techniques that are well known in the art. Five general methods for delivering a nucleic segment into cells have been described: (1) chemical methods (Graham and VanDerEb, 1973); (2) physical methods such as microinjection (Capecchi, 1980), electroporation (U.S. Patent No. 5,472,869; Tomes *et al.*, 1990; Wong and Neumann, 1982; Fromm *et al.*, 1985), microprojectile bombardment (Wang *et al.*, 1988; Vain *et al.*, 1990; U.S. Patent No. 5,874,265, specifically incorporated herein by reference in its entirety), "gene gun" (Hilber *et al.*, 1994; Yang *et al.*, 1990); (3) viral vectors (Clapp, 1993; Danos and Heard, 1992; Eglitis and Anderson,

1988); (4) receptor-mediated mechanisms (Curiel *et al.*, 1991; Wagner *et al.*, 1992); and (5) bacterial-mediated delivery such as *A. tumefaciens* transformation (Smith and Hood, 1995).

For example, a large number of cloning vectors comprising a replication system in *E. coli* and a marker that permits selection of the transformed cells are available for preparation for the insertion of foreign genes into higher organisms, including plants. The vectors comprise, for example, plasmids (such as pBR322, pUC series, M13mp series, pACYC184, *etc*), cosmids, phage, and/or phagemids and the like. Accordingly, the disclosed polynucleotides can be inserted into a given vector at a suitable restriction site. The resulting plasmid may be used, for example, to transform bacterial cells such as *E. coli* or *A. tumefaciens*. The bacterial cells are then cultivated in a suitable nutrient medium, harvested and lysed. The plasmid is recovered. Sequence analysis, restriction analysis, electrophoresis, and other biochemical-molecular biological methods are generally carried out as methods of analysis. After each manipulation, the DNA sequence used can be cleaved and joined to the next DNA sequence. Each plasmid sequence can be cloned in the same or other plasmids. Depending on the method of inserting desired genes into the plant, other DNA sequences may be necessary.

Methods for DNA transformation of plant cells include *Agrobacterium*-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs, injection into immature embryos and particle bombardment. Each of these methods has distinct advantages and disadvantages. Thus, one particular method of introducing genes into a particular plant strain may not necessarily be the most effective for another plant strain, but it is well known which methods are useful for a particular plant strain.

Suitable methods are believed to include virtually any method by which DNA can be introduced into a cell, such as by *Agrobacterium* infection, direct delivery of DNA such as, for example, by PEG-mediated transformation of protoplasts (Omirulleh *et al.*, 1993), by desiccation/inhibition-mediated DNA uptake, by electroporation, by agitation with silicon carbide fibers, by acceleration of DNA coated particles, *etc*. In certain embodiments, acceleration methods are preferred and include, for example, microprojectile bombardment and the like.

Technology for introduction of DNA into cells is well known to those of skill in the art, and described hereinbelow in detail. Likewise, a large number of techniques are available for inserting DNA into a plant host cell. Those techniques include transformation with T-DNA using *A. tumefaciens* or *A. rhizogenes* as transformation agent, fusion, injection, or

electroporation as well as other possible methods. If agrobacteria are used for the transformation, the DNA to be inserted has to be cloned into special plasmids, namely either into an intermediate vector or into a binary vector. The intermediate vectors can be integrated into the Ti or Ri plasmid by homologous recombination owing to sequences that are homologous to sequences in the T-DNA. The Ti or Ri plasmid also comprises the *vir* region necessary for the transfer of the T-DNA.

Intermediate vectors cannot replicate themselves in agrobacteria. The intermediate vector can be transferred into *A. tumefaciens* by means of a helper plasmid (conjugation). Binary vectors can replicate themselves both in *E. coli* and in agrobacteria. They comprise a selection marker gene and a linker or polylinker that are framed by the right and left T-DNA border regions. They can be transformed directly into agrobacteria (Holsters *et al.*, 1978). The agrobacterium used as host cell is to comprise a plasmid carrying a *vir* region. The *vir* region is necessary for the transfer of the T-DNA into the plant cell. Additional t-DNA may be contained. The bacterium so transformed is used for the transformation of plant cells. Plant explants can advantageously be cultivated with *A. tumefaciens* or *A. rhizogenes* for the transfer of the DNA into the plant cell. Whole plants can then be regenerated from the infected plant material (for example, pieces of leaf, segments of stalk, roots, but also protoplasts or suspension-cultivated cells) in a suitable medium, which may contain antibiotics or biocides for selection. The plants so obtained can then be tested for the presence of the inserted DNA. No special demands are made of the plasmids in the case of injection and electroporation. It is possible to use ordinary plasmids, such as, for example, pUC derivatives. If, for example, the Ti or Ri plasmid is used for the transformation of the plant cell, then at least the right border, but often the right and the left border of the Ti or Ri plasmid T-DNA, has to be joined as the flanking region of the genes to be inserted. The use of T-DNA for the transformation of plant cells has been intensively researched and sufficiently described in Eur. Pat. Appl. No. EP 120 516; Hockema (1985); An *et al.*, 1985, Herrera-Estrella *et al.*, (1983), Bevan *et al.*, (1983), and Klee *et al.*, (1985).

A particularly useful Ti plasmid cassette vector for transformation of dicotyledonous plants consists of the enhanced CaMV35S promoter (EN-35S) and the 3′ end including polyadenylation signals from a soybean gene encoding the α′-subunit of β-conglycinin. Between these two elements is a multilinker containing multiple restriction sites for the insertion of genes of interest.

The vector preferably contains a segment of pBR322 which provides an origin of replication in *E. coli* and a region for homologous recombination with the disarmed T-DNA in *Agrobacterium* strain ACO; the *oriV* region from the broad host range plasmid RK1; the streptomycin/spectinomycin resistance gene from Tn7; and a chimeric NPTII gene, containing the CaMV35S promoter and the nopaline synthase (NOS) 3' end, which provides kanamycin resistance in transformed plant cells.

Optionally, the enhanced CaMV35S promoter may be replaced with the 1.5 kb mannopine synthase (MAS) promoter (Velten *et al.*, 1984). After incorporation of a DNA construct into the vector, it is introduced into *A. tumefaciens* strain ACO that contains a disarmed Ti plasmid. Cointegrate Ti plasmid vectors are selected and subsequently may be used to transform a dicotyledonous plant.

A. tumefaciens ACO is a disarmed strain similar to pTiB6SE described by Fraley *et al.*, (1985). For construction of ACO the starting *Agrobacterium* strain was the strain A208 that contains a nopaline-type Ti plasmid. The Ti plasmid was disarmed in a manner similar to that described by Fraley *et al.*, (1985) so that essentially all of the native T-DNA was removed except for the left border and a few hundred base pairs of T-DNA inside the left border. The remainder of the T-DNA extending to a point just beyond the right border was replaced with a novel piece of DNA including (from left to right) a segment of pBR322, the *oriV* region from plasmid RK2, and the kanamycin resistance gene from Tn601. The pBR322 and *oriV* segments are similar to these segments and provide a region of homology for cointegrate formation.

Once the inserted DNA has been integrated in the genome, it is relatively stable there and, as a rule, does not come out again. It normally contains a selection marker that confers on the transformed plant cells resistance to a biocide or an antibiotic, such as kanamycin, G 418, bleomycin, hygromycin, or chloramphenicol, *inter alia*. The individually employed marker should accordingly permit the selection of transformed cells rather than cells that do not contain the inserted DNA.

4.1.1 ELECTROPORATION

The application of brief, high-voltage electric pulses to a variety of animal and plant cells leads to the formation of nanometer-sized pores in the plasma membrane. DNA is taken directly into the cell cytoplasm either through these pores or as a consequence of the redistribution of membrane components that accompanies closure of the pores. Electroporation can be extremely

efficient and can be used both for transient expression of clones genes and for establishment of cell lines that carry integrated copies of the gene of interest. Electroporation, in contrast to calcium phosphate-mediated transfection and protoplast fusion, frequently gives rise to cell lines that carry one, or at most a few, integrated copies of the foreign DNA.

The introduction of DNA by electroporation is well-known to those of skill in the art (see *e.g.*, U.S. Patent No. 5,324,253). In this method, certain cell wall-degrading enzymes, such as pectin-degrading enzymes, are employed to render the target recipient cells more susceptible to transformation by electroporation than untreated cells. Alternatively, recipient cells are made more susceptible to transformation, by mechanical wounding. To effect transformation by electroporation one may employ either friable tissues such as a suspension culture of cells (U.S. Patent No. 5,484,956; U.S. Patent No. 5,886,244), or embryogenic callus (U.S. Patent No. 5,405,765), or alternatively, one may transform immature embryos or other organized tissues directly. One would partially degrade the cell walls of the chosen cells by exposing them to pectin-degrading enzymes (pectolyases) or mechanically wounding in a controlled manner. Such cells would then be recipient to DNA transfer by electroporation, which may be carried out at this stage, and transformed cells then identified by a suitable selection or screening protocol dependent on the nature of the newly incorporated DNA.

4.1.2 MICROPROJECTILE BOMBARDMENT

A further advantageous method for delivering transforming DNA segments to plant cells is microprojectile bombardment. In this method, particles may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum, and the like.

An advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly stably transforming monocots, is that neither the isolation of protoplasts (Cristou *et al.*, 1988) nor the susceptibility to *Agrobacterium* infection is required. An illustrative embodiment of a method for delivering DNA into maize cells by acceleration is a Biolistics Particle Delivery System, which can be used to propel particles coated with DNA or cells through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with corn cells cultured in suspension. The screen disperses the particles so that they are not delivered to the recipient cells in large aggregates. It is believed that a screen intervening between the projectile apparatus and the cells to be bombarded reduces the size of projectiles aggregate and

may contribute to a higher frequency of transformation by reducing damage inflicted on the recipient cells by projectiles that are too large.

For the bombardment, cells in suspension are preferably concentrated on filters or solid culture medium. Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate. If desired, one or more screens are also positioned between the acceleration device and the cells to be bombarded. Through the use of techniques set forth herein one may obtain up to 1000 or more foci of cells transiently expressing a marker gene. The number of cells in a focus which express the exogenous gene product 48 h post-bombardment often range from 1 to 10 and average 1 to 3.

In bombardment transformation, one may optimize the prebombardment culturing conditions and the bombardment parameters to yield the maximum numbers of stable transformants. Both the physical and biological parameters for bombardment are important in this technology. Physical factors are those that involve manipulating the DNA/microprojectile precipitate or those that affect the flight and velocity of either the macro- or microprojectiles. Biological factors include all steps involved in manipulation of cells before and immediately after bombardment, the osmotic adjustment of target cells to help alleviate the trauma associated with bombardment, and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. It is believed that pre-bombardment manipulations are especially important for successful transformation of immature embryos.

Accordingly, it is contemplated that one may wish to adjust several of the bombardment parameters in small-scale studies to fully optimize the conditions. One may particularly wish to adjust physical parameters such as gap distance, flight distance, tissue distance, and helium pressure. One may also minimize the trauma reduction factors (TRFs) by modifying conditions which influence the physiological state of the recipient cells and which may therefore influence transformation and integration efficiencies. For example, the osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells may be adjusted for optimum transformation. The execution of other routine adjustments will be known to those of skill in the art in light of the present disclosure.

4.1.3 *AGROBACTERIUM*-MEDIATED TRANSFER

Agrobacterium-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for example, the methods described (Fraley *et al.*, 1985; Rogers *et al.*, 1988). Further, the integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences, and intervening DNA is usually inserted into the plant genome as described (Spielmann *et al.*, 1986; Jorgensen *et al.*, 1987).

Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as described (Klee *et al.*, 1985). Moreover, recent technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide-coding genes. The vectors described (Eichholtz *et al.*, 1987), have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes. In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

Agrobacterium-mediated transformation of leaf disks and other tissues such as cotyledons and hypocotyls appears to be limited to plants that *Agrobacterium* naturally infects. *Agrobacterium*-mediated transformation is most efficient in dicotyledonous plants. Few monocots appear to be natural hosts for *Agrobacterium*, although transgenic plants have been produced in asparagus using *Agrobacterium* vectors as described (Bytebier *et al.*, 1987). Therefore, commercially important cereal grains such as rice, corn, and wheat must usually be transformed using alternative methods (see e.g., U.S. Patent No. 5,610,042).

A transgenic plant formed using *Agrobacterium* transformation methods typically contains a single gene on one chromosome. Such transgenic plants can be referred to as being heterozygous for the added gene. However, inasmuch as use of the word "heterozygous" usually implies the presence of a complementary gene at the same locus of the second chromosome of a

pair of chromosomes, and there is no such gene in a plant containing one added gene as here, it is believed that a more accurate name for such a plant is an independent segregant, because the added, exogenous gene segregates independently during mitosis and meiosis.

More preferred is a transgenic plant that is homozygous for the added structural gene; *i.e.* a transgenic plant that contains two added genes, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) an independent segregant transgenic plant that contains a single added gene, germinating some of the seed produced and analyzing the resulting plants produced for enhanced carboxylase activity relative to a control (native, non-transgenic) or an independent segregant transgenic plant.

It is to be understood that two different transgenic plants can also be mated to produce offspring that contain two or more independently segregating added, exogenous genes. Selfing of appropriate progeny can produce plants that are homozygous for both added, exogenous genes that encode a polypeptide of interest. Backcrossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated.

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments (see, *e.g.*, Potrykus *et al.*, 1985; Lorz *et al.*, 1985; Fromm *et al.*, 1985; Uchimiya *et al.*, 1986; Callis *et al.*, 1987; Marcotte *et al.*, 1988).

Application of these systems to different plant strains depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts are described (Fujimura *et al.*, 1985; Toriyama *et al.*, 1986; Yamada *et al.*, 1986; Abdullah *et al.*, 1986).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil, 1988). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil, 1992).

Using that latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al.*, 1987; Klein *et al.*, 1988a; 1988b; McCabe *et al.*, 1988). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants.

4.2 EXPRESSION VECTORS

The present invention also provides an expression vector comprising at least one MMT-encoding gene-containing polynucleotide operably linked to an inducible promoter. Thus, in one embodiment an expression vector is an isolated and purified DNA molecule comprising an MMT coding region operably linked to a promoter that expresses the gene, which coding region is operatively linked to a transcription-terminating region, whereby the promoter drives the transcription of the coding region.

As used herein, the term "operatively linked" means that a promoter is connected to a nucleic acid region encoding functional RNA in such a way that the transcription of that functional RNA is controlled and regulated by that promoter. Means for operatively linking a promoter to a nucleic acid region encoding functional RNA are well known in the art.

The choice of which expression vector and ultimately to which promoter a polypeptide coding region is operatively linked depend directly on the functional properties desired, *e.g.*, the location and timing of protein expression, and the host cell to be transformed. These are well known limitations inherent in the art of constructing recombinant DNA molecules. However, a vector useful in practicing the present invention is capable of directing the expression of the functional RNA to which it is operatively linked.

RNA polymerase transcribes a coding DNA sequence through a site where polyadenylation occurs. Typically, DNA sequences located a few hundred base pairs downstream of the polyadenylation site serve to terminate transcription. Those DNA sequences are referred to herein as transcription-termination regions. Those regions are required for efficient polyadenylation of transcribed messenger RNA (mRNA).

A variety of methods have been developed to operatively link DNA to vectors *via* complementary cohesive termini or blunt ends. For instance, complementary homopolymer tracts can be added to the DNA segment to be inserted and to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

4.3 DNA SEGMENTS AS HYBRIDIZATION PROBES AND PRIMERS

In another aspect, DNA sequence information provided by the invention allows for the preparation of relatively short DNA (or RNA) sequences having the ability to specifically hybridize to gene sequences of the selected polynucleotides disclosed herein. The ability of such

nucleic acid probes to specifically hybridize to all or portions of one or more MMT-encoding genes lends them particular utility in a variety of embodiments. Most importantly, the probes may be used in a variety of assays for detecting the presence of complementary sequences in a given sample, and in the identification of new species or genera of MMT-encoding genes from a variety of host organisms.

In certain embodiments, it is advantageous to use oligonucleotide primers. The sequence of such primers is designed using a polynucleotide of the present invention for use in detecting, amplifying or mutating a defined segment of the disclosed MMT-encoding genes (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, and SEQ ID NO:5) from a sample using PCR™ technology. Segments of related MMT-encoding genes from other species, and particularly from other related plant species may also be amplified by PCR™ using such primers.

To provide certain of the advantages in accordance with the present invention, a preferred nucleic acid sequence employed for hybridization studies or assays includes sequences that are complementary to at least about 31 to 50 or so long nucleotide stretch of an MMT-encoding gene sequence. A size of at least 31 nucleotides in length helps to ensure that the fragment will be of sufficient length to form a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 31 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of about 31 to about 40 or 50 or so nucleotides, or even longer where desired. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, by application of nucleic acid reproduction technology, such as the PCR™ technology of U.S. Patent No. 4,683,195, and U.S. Patent No. 4,683,202, (each specifically incorporated herein by reference in its entirety), or by excising selected DNA fragments from recombinant plasmids containing appropriate inserts and suitable restriction sites.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template or where one seeks to isolate MMT-encoding gene sequences from related species, functional equivalents, or the like, less stringent hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ conditions such as about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-

hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

In addition to the use in directing the expression of functional RNA of the present invention, the nucleic acid sequences contemplated herein also have a variety of other uses. For example, they also have utility as probes or primers in nucleic acid hybridization embodiments. As such, it is contemplated that nucleic acid segments that comprise a sequence region that consists of at least a 14 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 14 nucleotide long contiguous DNA segment of one or more MMT-encoding genes will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 21, 22, 23, 24, *etc.*, 30, 31, 32, 33, 34, *etc.*, 40, 41, 42, 43, 44, *etc.*, 50, 51, 52, 53, 54, *etc.*, 100, 200, 500, 1000, 2000, 5000, 10000 *etc.* (including all intermediate lengths and up to and including full-length sequences will also be of use in certain embodiments.

While the ability of such nucleic acid probes to specifically hybridize to MMT-encoding gene sequences makes them ideal for use in detecting the presence of complementary sequences in a given sample, other uses are also envisioned, including the use of the sequence information for the preparation of mutant species primers, synthetic gene sequences, gene fusions, and/or primers for use in preparing other MMT-encoding genetic constructs.

The use of a hybridization probe of about 14 or so nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches of about 15, 16, 17, 18, 19, or 20 or more bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of about 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 or more contiguous nucleotides in length where desired.

Of course, fragments may also be obtained by other techniques such as, *e.g.*, by mechanical shearing or by restriction enzyme digestion. Small nucleic acid segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by

chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR™ technology of U.S. Patent No. 4,683,195 and U.S. Patent No. 4,683,202 (each incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of DNA fragments. Depending on the application envisioned, one may desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating particular DNA segments. Detection of DNA segments via hybridization is well known to those of skill in the art, and the teachings of U.S. Patent No. 4,965,188 and U.S. Patent No. 5,176,995 (each incorporated herein by reference) are exemplary of the methods of hybridization analyses. Teachings such as those found in the texts of Maloy *et al.*, 1994; Segal 1976; Prokop and Bajpai, 1991; and Kuby, 1994, are particularly relevant.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template or where one seeks to isolate an MMT-encoding gene from a related species, functional equivalents, or the like, less stringent hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ conditions such as about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In general, it is envisioned that the hybridization probes described herein will be useful both as reagents in solution hybridization as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to specific hybridization with selected probes under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria required (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, *etc.*). Following washing of the hybridized surface so as to remove nonspecifically bound probe molecules, specific hybridization is detected, or even quantitated, by means of the label.

4.4 Genetic Constructs Comprising MMT-Encoding Polynucleotide Sequences

An important embodiment of the present invention concerns genetic constructs, such as plasmids, vectors, viruses, cosmids, and the like that comprise one or more of the MMT-encoding polynucleotide sequences disclosed herein. Another important embodiment concerns genetic constructs that comprise one or more portions of an MMT-encoding polynucleotide or one or more regulatory regions of an MMT-encoding polynucleotide sequence. Particularly important genetic constructs include those used for the preparation of oligonucleotide probes, polynucleotide primers, recombinant proteins, peptides, or peptide epitopes, and those used for sequencing MMT-encoding gene sequences and homologous polynucleotides. Other important genetic constructs include transformation vectors, viruses and the like used for the introduction of the disclosed gene sequences into a host cell, plant tissue, or plant for the preparation of

transgenic plants or transformed host cells that express the MMT polypeptide when cultured under appropriate conditions.

A variety of genetic compositions may be used for preparation and delivery of the disclosed MMT genetic constructs to selected recipient host cells, and particularly to selected plant host cells or tissues to ultimately produce transformed plants and plant cell lines in accordance with the present invention. For example, polynucleotides in the form of vectors and plasmids, or linear nucleic acid fragments, in some instances containing only the particular polynucleotide to be expressed in the cell, tissue or animal, and the like, may be employed.

Vectors, plasmids, phagemids, cosmids, viral vectors, shuttle vectors, baculovirus vectors, BACs (bacterial artificial chromosomes), PACs (plant artificial chromosomes), YACs (yeast artificial chromosomes) and DNA segments for use in transforming cells with a nucleic acid construct of interest, are well known to those of skill in the microbiological and plant molecular biology arts. Typically such constructs generally comprise at least one promoter or other regulatory region that is operably linked to at least one or more of the novel polynucleotides disclosed herein. These polynucleotide constructs may contain a cDNA, or one or more genes which one desires to introduce into a particular cell, cell line, tissue, or other suitable organism. Such polynucleotide constructs may also optionally include one or more structures such as inducible, constitutive, or tissue-specific promoters, one or more enhancers or enhancer elements, one or more polylinkers or multiple cloning sites, or one or more regulatory sequences as may be desired. The polynucleotide segment or gene chosen for cellular introduction may encode the entire MMT protein such that the protein may be expressed in the resultant recombinant cells, or, alternatively, the nucleic acid constructs may contain portions of the coding region, or an MMT gene regulatory region alone or in combination with other gene sequences, or may even comprise one or more antisense constructs, or ribozyme-encoding regions.

4.5 Methods for Preparing Mutagenized Polynucleotides

In certain circumstances, it may be desirable to modify or alter one or more nucleotides in one or more of the sequences disclosed herein for the purpose of altering or changing the transcriptional activity or other property of the sequence region. In general, the means and methods for mutagenizing a polynucleotide are well known to those of skill in the art. Modifications to such polynucleotides may be made by random, or site-specific mutagenesis

procedures. The selected polynucleotide may be modified by altering its structure through the addition or deletion of one or more nucleotides within the sequence, or may be modified by the addition of a cloning site, a polylinkers region, or by the preparation of a gene fusion or a protein fusion encoding polynucleotide. Means for preparing mutagenized polynucleotides are exemplified in a number of U.S. patents and in the scientific literature. For example, U.S. Patent No. 6,023,013 (specifically incorporated herein by reference in its entirety) provides a variety of methods for preparing mutagenized polynucleotides.

4.6 Expression of Transgenes in Plants

In many instances, the level of transcription of a particular transgene in a given host cell is not always indicative of the amount of protein being produced in the transformed host cell. This is often due to post-transcriptional processes, such as splicing, polyadenylation, appropriate translation initiation, and RNA stability that affect the ability of a transcript to produce protein. Such factors may also affect the stability and amount of mRNA produced from the given transgene. As such, it is often desirable to alter the post-translational events through particular molecular biology techniques. The inventors contemplate that in certain instances it may be desirable to alter the transcription and/or expression of the MMT-encoding gene constructs of the present invention to increase, decrease, or otherwise regulate or control these constructs in particular host cells and/or transgenic plants.

4.6.1 Efficient Initiation of Protein Translation

The 5′-untranslated leader (5′-UTL) sequence of eukaryotic mRNA plays a major role in translational efficiency. Many early chimeric transgenes using a viral promoter used an arbitrary length of viral sequence after the transcription initiation site and fused this to the AUG of the coding region. More recently studies have shown that the 5′-UTL sequence and the sequences directly surrounding the AUG can have a large effect in translational efficiency in host cells and particularly certain plant species and that this effect can be different depending on the particular cells or tissues in which the message is expressed.

In most eukaryotic mRNAs, the point of translational initiation occurs at the AUG codon closest to the 5′ cap of the transcript. Comparison of plant mRNA sequences and site directed mutagenesis experiments have demonstrated the existence of a consensus sequence surrounding the initiation codon in plants (Joshi, 1987; Lutcke *et al.*, 1987). However, consensus sequences

will be apparent amongst individual plant species. For example, a compilation of sequences surrounding the initiation codon from 85 maize genes yields a consensus of 5′-(C/G)AUGGCG-3′ (Luehrsen *et al.*, 1994). In tobacco protoplasts, transgenes encoding β-glucuronidase (GUS) and bacterial chitinase showed a 4-fold and an eight-fold increase in expression, respectively, when the native sequences of these genes were changed to encode 5′-ACCAUGG-3′ (Gallie *et al.*, 1987b; Jones *et al.*, 1988).

When producing chimeric transgenes (*i.e.* transgenes comprising DNA segments from different sources operably linked together), often the 5′-UTL of plant viruses is used. The alfalfa mosaic virus (AMV) coat protein and brome mosaic virus (BMV) coat protein 5′-UTLs have been shown to enhance mRNA translation 8-fold in electroporated tobacco protoplasts (Gallie *et al.*, 1987a; 1987b). A 67-nucleotide derivative (Ω) of the 5′-UTL of tobacco mosaic virus RNA (TMV) fused to the chloramphenicol acetyltransferase (CAT) gene and GUS gene has been shown to enhance translation of reporter genes *in vitro* (Gallie *et al.*, 1987a; 1987b; Sleat *et al.*, 1987; Sleat *et al.*, 1988). Electroporation of tobacco mesophyll protoplasts with transcripts containing the TMV leader fused to reporter genes CAT, GUS, and LUC produced a 33-, 21-, and 36-fold level of enhancement, respectively (Gallie *et al.*, 1987a; 1987b; Gallie *et al.*, 1991). Also in tobacco, an 83-nt 5′-UTL of potato virus X RNA was shown to enhance expression of the neomycin phosphotransferese II (*Npt*II) 4-fold (Poogin and Skryabin, 1992).

The effect of a 5′-UTL may be different depending on the plant, particularly between dicots and monocots. The TMV 5′-UTL has been shown to be more effective in tobacco protoplasts (Gallie *et al.*, 1989) than in maize protoplasts (Gallie and Young, 1994). Also, the 5′-UTLs from TMV-Ω (Gallie *et al.*, 1988), AMV-coat (Gehrke *et al.*, 1983; Jobling and Gehrke, 1987), TMV-coat (Goelet *et al.*, 1982), and BMV-coat (French *et al.*, 1986) worked poorly in maize and inhibited expression of a luciferase gene in maize relative to its native leader (Koziel *et al.*, 1996). However, the 5′-UTLs from the cauliflower mosaic virus (CaMV) 35S transcript and the maize genes glutelin (Boronat *et al.*, 1986), PEP-carboxylase (Hudspeth and Grula, 1989) and ribulose biphosphate carboxylase showed a considerable increase in expression of the luciferase gene in maize relative to its native leader (Koziel *et al.*, 1996).

These 5′-UTLs had different effects in tobacco. In contrast to maize, the TMV Ω 5′-UTL and the AMV coat protein 5′-UTL enhanced expression in tobacco, whereas the glutelin, maize PEP-carboxylase and maize ribulose-1,5-bisphosphate carboxylase 5′-UTLs did not show enhancement relative to the native luciferase 5′-UTL (Koziel *et al.*, 1996). Only the CaMV 35S

5′-UTL region enhanced luciferase expression in both maize and tobacco (Koziel *et al.*, 1996). Furthermore, the TMV and BMV coat protein 5′-UTLs were inhibitory in both maize and tobacco protoplasts (Koziel *et al.*, 1996).

4.6.2 Use of Introns to Increase Expression

Including one or more introns in the transcribed portion of a gene has been found to increase heterologous gene expression in a variety of plant systems (Callis *et al.*, 1987; Maas *et al.*, 1991; Mascerenhas *et al.*, 1990; McElroy *et al.*, 1990; Vasil *et al.*, 1989), although not all introns produce a stimulatory effect and the degree of stimulation varies. The enhancing effect of introns appears to be more apparent in monocots than in dicots. Tanaka *et al.*, (1990) has shown that use of the catalase intron 1 isolated from castor beans increases gene expression in rice. Likewise, the first intron of the alcohol dehydrogenase 1 (*Adh1*) has been shown to increase expression of a genomic clone of *Adh1* comprising the endogenous promoter in transformed maize cells (Callis *et al.*, 1987; Dennis *et al.*, 1984). Other introns that are also able to increase expression of transgenes which contain them include introns 2 and 6 of *Adh1* (Luehrsen and Walbot, 1991), the catalase intron (Tanaka *et al.*, 1990), intron 1 of the maize bronze 1 gene (Callis *et al.*, 1987), the maize sucrose synthase intron 1 (Vasil *et al.*, 1989), intron 3 of the rice actin gene (Luehrsen and Walbot, 1991), rice actin intron 1 (McElroy *et al.*, 1990), and the heat shock protein HSP70 (U.S. Patent No. 5,859,347, specifically incorporated herein by reference in its entirety). Similar results may also be obtained using sequences from certain exons, for example, the maize ubiquitin exon 1 (Christensen *et al.*, 1992).

Generally, to achieve optimal expression, the selected intron(s) should be present in the 5′ transcriptional unit in the correct orientation with respect to the splice junction sequences (Callis *et al.*, 1987; Maas *et al.*, 1991; Mascerenhas *et al.*, 1990; Oard *et al.*, 1989; Tanaka *et al.*, 1990; Vasil *et al.*, 1989). Intron 9 of *Adh1* has been shown to increase expression of a heterologous gene when placed 3′ (or downstream of) the gene of interest (Callis *et al.*, 1987).

4.6.3 Use of Synthetic Genes to Increase Gene Expression

When introducing a prokaryotic gene into a eukaryotic host, or when expressing a eukaryotic gene in a non-native host, the sequence of the gene must often be altered or modified to allow efficient translation of the transcript(s) derived from the gene. Significant experience in using synthetic genes to increase expression of a desired protein has been achieved in the

expression of *B. thuringiensis*-derived genes in plants. Native *B. thuringiensis* genes are expressed only at low levels in dicots and not at all in monocots (Koziel *et al.*, 1996). Codon usage in the native genes is considerably different from that found in typical plant genes, which have a higher G+C content. Strategies to increase expression of these genes in plants generally alter the overall G+C content of the genes. For example, synthetic *B. thuringiensis* δ-endotoxin encoding genes have resulted in significant improvements in expression of the δ-endotoxins in various crops including cotton (Perlak *et al.*, 1990; Wilson *et al.*, 1992), tomato (Perlak *et al.*, 1991), potato (Perlak *et al.*, 1993), rice (Cheng *et al.*, 1998), and maize (Koziel *et al.*, 1993).

In a similar fashion the inventors contemplate that the MMT-encoding sequences of the present invention, may in certain circumstances, be altered to increase or decrease the expression of the genes in particular eukaryotic host cells and/or transgenic plants that comprise such constructs. Using molecular biology techniques that are well known to those of skill in the art, one may alter the coding or non coding sequences of the particular MMT gene(s) to optimize or facilitate its expression in transformed plant cells at suitable levels.

4.6.4 CHLOROPLAST SEQUESTERING AND TARGETING

Methods of targeting proteins to the chloroplast have been developed utilizing the pea chloroplast transit peptide, to target the enzymes of the polyhydroxybutyrate synthesis pathway to the chloroplast (Nawrath *et al.*, 1994). Also, this technique negated the necessity of modification of the coding region other than to add an appropriate targeting sequence.

U.S. Patent No. 5,576,198 (specifically incorporated herein by reference) discloses compositions and methods useful for genetic engineering of plant cells to provide a method of controlling the timing or tissue pattern of expression of foreign DNA sequences inserted into the plant plastid genome. Constructs include those for nuclear transformation that provide for expression of a viral single subunit RNA polymerase in plant tissues, and targeting of the expressed polymerase protein into plant cell plastids. Also included are plastid expression constructs comprising a viral gene promoter region which is specific to the RNA polymerase expressed from the nuclear expression constructs described above and a heterologous gene of interest to be expressed in the transformed plastid cells.

4.6.5 Effects of 3′ Regions on Transgene Expression

The 3′-end regions of transgenes have been found to have a large effect on transgene expression in plants (Ingelbrecht *et al.*, 1989). In this study, different 3′ ends were operably linked to the neomycin phosphotransferase II (*NptII*) reporter gene and expressed in transgenic tobacco. The different 3′ ends used were obtained from the octopine synthase gene, the 2S seed protein from *Arabidopsis*, the small subunit of *rbcS* from *Arabidopsis*, extension form carrot, and chalcone synthase from *Antirrhinum*. In stable tobacco transformants, there was about a 60-fold difference between the best-expressing construct (small subunit *rbcS* 3′ end) and the lowest expressing construct (chalcone synthase 3′ end).

TABLE 1

PLANT PROMOTERS

Promoter	Reference[a]
Viral	
Figwort Mosaic Virus (FMV)	U.S. Patent No. 5,378,619
Cauliflower Mosaic Virus (CaMV)	U.S. Patent No. 5,530,196
	U.S. Patent No. 5,097,025
	U.S. Patent No. 5,110,732
Plant	
Elongation Factor	U.S. Patent No. 5,177,011
Tomato Polygalacturonase	U.S. Patent No. 5,442,052
Arabidopsis Histone H4	U.S. Patent No. 5,491,288
Phaseolin	U.S. Patent No. 5,504,200
Group 2	U.S. Patent No. 5,608,144
Ubiquitin	U.S. Patent No. 5,614,399
P119	U.S. Patent No. 5,633,440
α-amylase	U.S. Patent No. 5,712,112
Wheat starch branching enzyme	U.S. Patent No. 5,866,793
Osmotin	U.S. Patent No. 5,874,626
Viral enhancer/Plant promoter	
CaMV 35S enhancer/mannopine synthase promoter	U.S. Patent No. 5,106,739

[a]Each reference is specifically incorporated herein by reference in its entirety.

TABLE 2

TISSUE SPECIFIC PLANT PROMOTERS

Tissue Specific Promoter	Tissue(s)	Reference[a]
Blec	Epidermis	U.S. Patent No. 5,646,333
Malate synthase	Seeds; seedlings	U.S. Patent No. 5,689,040
Isocitrate lyase	Seeds; seedlings	U.S. Patent No. 5,689,040
Patatin	Tuber	U.S. Patent No. 5,436,393
ZRP2	Root	U.S. Patent No. 5,633,363
ZRP2(2.0)	Root	U.S. Patent No. 5,633,363
ZRP2(1.0)	Root	U.S. Patent No. 5,633,363
RB7	Root	U.S. Patent No. 5,459,252
	Root	U.S. Patent No. 5,401,836
	Fruit	U.S. Patent No. 4,943,674
	Meristem	U.S. Patent No. 5,589,583
	Guard cell	U.S. Patent No. 5,538,879
	Stamen	U.S. Patent No. 5,589,610
SodA1	Pollen; middle layer; stomium of anthers	Van Camp *et al.*, 1996
SodA2	Vascular bundles; stomata; axillary buds; pericycle; stomium; pollen	Van Camp *et al.*, 1996
CHS15	Flowers; root tips	Faktor *et al.*, 1996
Psam-1	Phloem tissue; cortex; root tips	Vander *et al.*, 1996
ACT11	Elongating tissues and organs; pollen; ovules	Huang *et al.*, 1997
ZmGBS	Pollen; endosperm	Russell and Fromm, 1997
zmZ27	Endosperm	Russell and Fromm, 1997
OsAGP	Endosperm	Russell and Fromm, 1997
osGT1	Endosperm	Russell and Fromm, 1997
RolC	Phloem tissue; bundle sheath;	Graham *et al.*, 1997

Tissue Specific Promoter	Tissue(s)	Reference[a]
	vascular parenchyma	
Sh	Phloem tissue	Graham *et al.*, 1997
CMd	Endosperm	Grosset *et al.*, 1997
Bnm1	Pollen	Treacy *et al.*, 1997
rice tungro bacilliform virus	Phloem	Yin *et al.*, 1997a; 1997b
S2-RNase	Pollen	Ficker *et al.*, 1998
LeB4	Seeds	Baumlein *et al.*, 1991
gf-2.8	Seeds; seedlings	Berna and Bernier, 1997

[a]Each reference is specifically incorporated herein by reference in its entirety.

The ability to express genes in a tissue specific manner in plants has led to the production of male and female sterile plants. Generally, the production of male sterile plants involves the use of anther-specific promoters operably linked to heterologous genes that disrupt pollen formation (U.S. Patent Nos. 5,689,051; 5,689,049; 5,659,124, each specifically incorporated herein by reference). U.S. Patent No. 5,633,441 (specifically incorporated herein by reference) discloses a method of producing plants with female genetic sterility. The method comprises the use of style-cell, stigma-cell, or style- and stigma-cell specific promoters that express polypeptides that, when produced in the cells of the plant kill or significantly disturbs the metabolism, functioning or development of the cells.

TABLE 3

INDUCIBLE PLANT PROMOTERS

Promoter	Reference[a]
Heat shock promoter	U.S. Patent No. 5,447,858
Em	U.S. Patent No. 5,139,954
Adh1	Kyozuka *et al.*, 1991
HMG2	U.S. Patent No. 5,689,056
Cinnamyl alcohol dehydrogenase	U.S. Patent No. 5,633,439
Asparagine synthase	U.S. Patent No. 5,595,896
GST-II-27	U.S. Patent No. 5,589,614

[a]Each reference is specifically incorporated herein by reference in its entirety.

4.7 GENE EXPRESSION IN PLANTS

Although great progress has been made in recent years with respect to preparation of transgenic plants that express MMT-polypeptides and DNA binding proteins such as the *lac* operator, the results of expressing heterologous genes in particular plant species are often disappointing. Unlike microbial genetics, little was known by early plant geneticists about the factors that affected heterologous expression of foreign genes in plants. In recent years, however, several potential factors have been implicated as responsible in varying degrees for the level of protein expression from a particular coding sequence. For example, scientists now know that maintaining a significant level of a particular mRNA in the cell is indeed a critical factor. Unfortunately, the causes for low steady state levels of mRNA encoding foreign proteins are many. First, full length RNA synthesis may not occur at a high frequency. This could, for example, be caused by the premature termination of RNA during transcription or due to unexpected mRNA processing during transcription. Second, full length RNA may be produced in the plant cell, but then processed (splicing, polyA addition) in the nucleus in a fashion that creates a nonfunctional mRNA. If the RNA is not properly synthesized, terminated and polyadenylated, it cannot move to the cytoplasm for translation. Similarly, in the cytoplasm, if mRNAs have reduced half-lives (which are determined by their primary or secondary sequence) insufficient protein product will be produced. In addition, there is an effect, whose magnitude is uncertain, of translational efficiency on mRNA half-life. In addition, every RNA molecule folds

into a particular structure, or perhaps family of structures, which is determined by its sequence. The particular structure of any RNA might lead to greater or lesser stability in the cytoplasm. Structure *per se* is probably also a determinant of mRNA processing in the nucleus. Unfortunately, it is impossible to predict, and nearly impossible to determine, the structure of any RNA (except for tRNA) *in vitro* or *in vivo*. However, it is likely that dramatically changing the sequence of an RNA will have a large effect on its folded structure. It is likely that structure *per se* or particular structural features also have a role in determining RNA stability.

To overcome these limitations in foreign gene expression, researchers have identified particular sequences and signals in RNAs that have the potential for having a specific effect on RNA stability. In certain embodiments of the invention, therefore, there is a desire to optimize expression of the disclosed nucleic acid segments *in planta*. One particular method of doing so is by alteration of the bacterial gene to remove sequences or motifs that decrease expression in a transformed plant cell. The process of engineering a coding sequence for optimal expression *in planta* is often referred to as "plantizing" a DNA sequence.

Particularly problematic sequences are those that are A+T rich. Unfortunately, since many bacterial species have genomes that are rich in A+T sequences, native bacterial gene sequences must often be modified for optimal expression in eukaryotes, and particularly in a transformed plant. Many short-lived mRNAs have A+T rich 3′ untranslated regions, and these regions often have the ATTTA sequence, sometimes present in multiple copies or as multimers (*e.g.*, ATTTATTTA...). Shaw and Kamen showed that the transfer of the 3′ end of an unstable mRNA to a stable RNA (globin or VA1) decreased the stable RNA's half-life dramatically (Shaw and Kamen, 1986; 1987). They further showed that a pentamer of ATTTA had a profound destabilizing effect on a stable message, and that this signal could exert its effect whether it was located at the 3′-end or within the coding sequence. However, the number of ATTTA sequences and/or the sequence context in which they occur also appears to be important in determining whether they function as a destabilizing sequence.

The addition of a polyadenylation sequence to the 3′-end is common to most eukaryotic mRNAs, both plant and animal. The currently accepted view of polyA addition is that the nascent transcript extends beyond the mature 3′-terminus. Contained within this transcript are signals for polyadenylation and proper 3′-end formation. This processing at the 3′-end involves cleavage of the mRNA and addition of polyA to the mature 3′-end. By searching for consensus sequences near the polyA tract in both plant and animal mRNAs, it has been

possible to identify consensus sequences that apparently are involved in polyA addition and 3'-end cleavage. The same consensus sequences seem to be important to both of these processes. These signals are typically a variation on the sequence AATAAA. In animal cells, some variants of this sequence that are functional have been identified; in plant cells there seems to be an extended range of functional sequences (Wickens and Stephenson, 1984; Dean *et al.*, 1986). Because all of these consensus sequences are variations on AATAAA, they all are A+T rich sequences. This sequence is typically found 15 to 20 bp before the polyA tract in a mature mRNA. Studies in animal cells indicate that this sequence is involved in both polyA addition and 3'-maturation. Site-directed mutation in this sequence may disrupt these functions (Conway and Wickens, 1988; Wickens *et al.*, 1987). However, it has also been observed that sequences up to 50 to 100 bp 3' to the putative polyA signal are also required; *i.e.*, a gene that has a normal AATAAA but has been replaced or disrupted downstream does not get properly polyadenylated (Gil and Proudfoot, 1984; Sadofsky and Alwine, 1984; McDevitt *et al.*, 1984). That is, the polyA signal itself is not sufficient for complete and proper processing. It is not yet known what specific downstream sequences are required in addition to the polyA signal, or if there is a specific sequence that has this function. Therefore, sequence analysis can only identify potential polyA signals.

In naturally occurring mRNAs that are normally polyadenylated, it has been observed that disruption of this process, either by altering the polyA signal or other sequences in the mRNA, profound effects can be obtained in the level of functional mRNA. This has been observed in several naturally occurring mRNAs, with results that are gene-specific so far.

It has been shown that in natural mRNAs proper polyadenylation is important in mRNA accumulation, and that disruption of this process can affect mRNA levels significantly. However, insufficient knowledge exists to predict the effect of changes in a normal gene. In a heterologous gene, it is even harder to predict the consequences. However, it is possible that the putative sites identified are dysfunctional. That is, these sites may not act as proper polyA sites, but instead function as aberrant sites that give rise to unstable mRNAs.

In animal cell systems, AATAAA is by far the most common signal identified in mRNAs upstream of the polyA, but at least four variants have also been found (Wickens and Stephenson, 1984). In plants, not nearly so much analysis has been done, but it is clear that multiple sequences similar to AATAAA can be used. The plant sites in Table 4 called major or minor refer only to the study of Dean *et al.*, (1986) which analyzed only three types of plant

gene. The designation of polyadenylation sites as major or minor refers only to the frequency of their occurrence as functional sites in naturally occurring genes that have been analyzed. In the case of plants this is a very limited database. It is hard to predict with any certainty that a site designated major or minor is more or less likely to function partially or completely when found in a heterologous gene such as those encoding the MMT polypeptides of the present invention.

TABLE 4

POLYADENYLATION SITES IN PLANT GENES

PA	AATAAA	Major consensus site
P1A	AATAAT	Major plant site
P2A	AACCAA	Minor plant site
P3A	ATATAA	"
P4A	AATCAA	"
P5A	ATACTA	"
P6A	ATAAAA	"
P7A	ATGAAA	"
P8A	AAGCAT	"
P9A	ATTAAT	"
P10A	ATACAT	"
P11A	AAAATA	"
P12A	ATTAAA	Minor animal site
P13A	AATTAA	"
P14A	AATACA	"
P15A	CATAAA	"

The present invention provides a method for preparing synthetic MMT-encoding genes that express their polypeptide product at sufficiently high levels in a heterologous transformed plant, so as to alter the biosynthesis of sulfur-containing compounds in the transformed plant.

4.8 SYNTHETIC OLIGONUCLEOTIDES FOR MUTAGENESIS

When oligonucleotides are used in the mutagenesis, it is desirable to maintain the proper amino acid sequence and reading frame, without introducing common restriction sites such as *Bgl*II, *Hin*dIII, *Sac*I, *Kpn*I, *Eco*RI, *Nco*I, *Pst*I and *Sal*I into the modified gene. These restriction sites are found in poly-linker insertion sites of many cloning vectors. Of course, the introduction of new polyadenylation signals, ATTTA sequences or consecutive stretches of more than five A+T or G+C, should also be avoided. The preferred size for the oligonucleotides is about 40 to about 50 bases, but fragments ranging from about 18 to about 100 bases have been utilized. In most cases, a minimum of about 5 to about 8 base pairs of homology to the template DNA on both ends of the synthesized fragment are maintained to insure proper hybridization of the primer to the template. The oligonucleotides should avoid sequences longer than five base pairs A+T or G+C. Codons used in the replacement of wild-type codons should preferably avoid the TA or CG doublet wherever possible. Codons are selected from a plant preferred codon table (such as Table 5 below) so as to avoid codons which are rarely found in plant genomes, and efforts should be made to select codons to preferably adjust the G+C content to about 50%.

TABLE 5

PREFERRED CODON USAGE IN PLANTS

Amino Acid	Codon	Percent Usage in Plants
ARG	CGA	7
	CGC	11
	CGG	5
	CGU	25
	AGA	29
	AGG	23
LEU	CUA	8
	CUC	20
	CUG	10
	CUU	28

Amino Acid	Codon	Percent Usage in Plants
	UUA	5
	UUG	30
SER	UCA	14
	UCC	26
	UCG	3
	UCU	21
	AGC	21
	AGU	15
THR	ACA	21
	ACC	41
	ACG	7
	ACU	31
PRO	CCA	45
	CCC	19
	CCG	9
	CCU	26
ALA	GCA	23
	GCC	32
	GCG	3
	GCU	41
GLY	GGA	32
	GGC	20
	GGG	11
	GGU	37

Amino Acid	Codon	Percent Usage in Plants
ILE	AUA	12
	AUC	45
	AUU	43
VAL	GUA	9
	GUC	20
	GUG	28
	GUU	43
LYS	AAA	36
	AAG	64
ASN	AAC	72
	AAU	28
GLN	CAA	64
	CAG	36
HIS	CAC	65
	CAU	35
GLU	GAA	48
	GAG	52
ASP	GAC	48
	GAU	52
TYR	UAC	68
	UAU	32

Amino Acid	Codon	Percent Usage in Plants
CYS	UGC	78
	UGU	22
PHE	UUC	56
	UUU	44
MET	AUG	100
TRP	UGG	100

Regions with many consecutive A+T bases or G+C bases are predicted to have a higher likelihood to form hairpin structures due to self-complementarity. Disruption of these regions by the insertion of heterogeneous base pairs is preferred and should reduce the likelihood of the formation of self-complementary secondary structures such as hairpins which are known in some organisms to inhibit transcription (transcriptional terminators) and translation (attenuators).

Alternatively, a completely synthetic gene for a given amino acid sequence can be prepared, with regions of five or more consecutive A+T or G+C nucleotides being avoided. Codons are selected avoiding the TA and CG doublets in codons whenever possible. Codon usage can be normalized against a plant preferred codon usage table (such as Table 5) and the G+C content preferably adjusted to about 50%. The resulting sequence should be examined to ensure that there are minimal putative plant polyadenylation signals and ATTTA sequences. Restriction sites found in commonly used cloning vectors are also preferably avoided. However, placement of several unique restriction sites throughout the gene is useful for analysis of gene expression or construction of gene variants.

4.9 "PLANTIZED" GENE CONSTRUCTS

The expression of a plant gene that exists in double-stranded DNA form involves transcription of messenger RNA (mRNA) from one strand of the DNA by RNA polymerase enzyme, and the subsequent processing of the mRNA primary transcript inside the nucleus. This processing involves a 3′ non-translated region that adds polyadenylated nucleotides to the 3′ end of the RNA. Transcription of DNA into mRNA is regulated by a region of DNA usually referred to as the "promoter." The promoter region contains a sequence of bases that signals RNA polymerase to associate with the DNA and to initiate the transcription of mRNA using one of the DNA strands as a template to make a corresponding strand of RNA.

A number of promoters that are active in plant cells have been described in the literature. These include the nopaline synthase (NOS) and octopine synthase (OCS) promoters (which are carried on tumor-inducing plasmids of *A. tumefaciens*), the Cauliflower Mosaic Virus (CaMV) 19S and 35S promoters, the light-inducible promoter from the small subunit of ribulose bis-phosphate carboxylase (ssRUBISCO, a very abundant plant polypeptide) and the mannopine synthase (MAS) promoter (Velten *et al.*, 1984 and Velten and Schell, 1985). All of these promoters have been used to create various types of DNA constructs that have been expressed in plants (see *e.g.*, Int. Pat. Appl. Publ. No. WO 84/02913).

Promoters that are known or are found to cause transcription of RNA in plant cells can be used in the present invention. Such promoters may be obtained from plants or plant viruses and include, but are not limited to, the CaMV35S promoter and promoters isolated from plant genes such as ssRUBISCO genes. As described below, it is preferred that the particular promoter selected should be capable of causing sufficient expression to result in the production of an effective amount of protein.

The promoters used in the DNA constructs (*i.e.* chimeric plant genes) of the present invention may be modified, if desired, to affect their control characteristics. For example, the CaMV35S promoter may be ligated to the portion of the ssRUBISCO gene that represses the expression of ssRUBISCO in the absence of light, to create a promoter which is active in leaves but not in roots. The resulting chimeric promoter may be used as described herein. For purposes of this description, the phrase "CaMV35S" promoter thus includes variations of CaMV35S promoter, *e.g.*, promoters derived by means of ligation with operator regions, random or controlled mutagenesis, *etc.* Furthermore, the promoters may be altered to contain multiple "enhancer sequences" to assist in elevating gene expression.

The RNA produced by a DNA construct of the present invention also contains a 5′ non-translated leader sequence. This sequence can be derived from the promoter selected to express the gene, and can be specifically modified so as to increase translation of the mRNA. The 5′ non-translated regions can also be obtained from viral RNAs, from suitable eukaryotic genes, or from a synthetic gene sequence. The present invention is not limited to constructs, as presented in the following examples. Rather, the non-translated leader sequence can be part of the 5′ end of the non-translated region of the coding sequence for the virus coat protein, or part of the promoter sequence, or can be derived from an unrelated promoter or coding sequence. In any case, it is preferred that the sequence flanking the initiation site conform to the translational consensus sequence rules for enhanced translation initiation reported by Kozak (1984).

The DNA constructs of the present invention may also contain one or more modified or fully synthetic structural coding sequences which have been changed to enhance the performance of the gene in a particular species of plant. The structural genes of the present invention may optionally encode a fusion protein comprising an amino-terminal chloroplast transit peptide or secretory signal sequence.

The DNA construct also contains a 3′ non-translated region. The 3′ non-translated regions contain a polyadenylation signal which functions in plants to cause the addition of polyadenylate nucleotides to the 3′ end of the viral RNA. Examples of suitable 3′ regions are (1) the 3′ transcribed, non-translated regions containing the polyadenylation signal of *Agrobacterium* tumor-inducing (Ti) plasmid genes, such as the nopaline synthase (NOS) gene, and (2) plant genes like the soybean storage protein (7S) genes and the small subunit of the RuBP carboxylase (E9) gene.

4.10 Methods for Altering MMT Activity in Transgenic Plants

By transforming a suitable host cell, such as a plant cell, with a recombinant MMT-encoding polynucleotide segment, the expression of the MMT-encoding polynucleotide under the control of an inducible promoter can result in the formation of transgenic plants in which the alteration of MMT activity and subsequent modulation of sulfur-containing compound biosynthesis may be achieved.

By way of example, one may utilize an expression vector containing a coding region for an MMT polynucleotide and an appropriate selectable marker to transform a suspension of embryonic plant cells, such as wheat or corn cells using a method such as particle bombardment

(Maddock *et al.*, 1991; Vasil *et al.*, 1992) to deliver the DNA coated on microprojectiles into the recipient cells. Transgenic plants are then regenerated from transformed embryonic calli that express the encoded polypeptide.

The formation of transgenic plants may also be accomplished using other methods of cell transformation that are known in the art such as *Agrobacterium*-mediated DNA transfer (Fraley *et al.*, 1983). Alternatively, DNA can be introduced into plants by direct DNA transfer into pollen (U.S. Patent No. 5,629,183; Zhou *et al.*, 1983; Hess, 1987; Luo *et al.*, 1988), by injection of the DNA into reproductive organs of a plant (Pena *et al.*, 1987), or by direct injection of DNA into the cells of immature embryos followed by the rehydration of desiccated embryos (Neuhaus *et al.*, 1987; Benbrook *et al.*, 1986).

Methods for the regeneration, development, and cultivation of plants from single plant protoplast transformants or from various transformed explants are well known in the art (Weissbach and Weissbach, 1988). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil.

The development or regeneration of plants containing the foreign, exogenous gene that encodes a polypeptide of interest introduced by *Agrobacterium* from leaf explants can be achieved by methods well known in the art such as described (Horsch *et al.*, 1985). In this procedure, transformants are cultured in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant strain being transformed as described (Fraley *et al.*, 1983).

This procedure typically produces shoots within two to four months and those shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil or other media to allow the production of roots. These procedures vary depending upon the particular plant strain employed, such variations being well known in the art.

Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants, as discussed before. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important, preferably inbred lines. Conversely, pollen from

plants of those important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

Such plants can form germ cells and transmit the transformed trait(s) to progeny plants. Likewise, transgenic plants can be grown in the normal manner and crossed with plants that have the same transformed hereditary factors or other hereditary factors. The resulting hybrid individuals have the corresponding phenotypic properties. A transgenic plant of this invention thus has an increased amount of a coding region that encodes the MMT polypeptide of interest. A preferred transgenic plant is an independent segregant and can transmit that gene and its activity to its progeny. A more preferred transgenic plant is homozygous for that gene, and transmits that gene to each of its offspring on sexual mating.

Seed from a transgenic plant may be grown in the field or greenhouse, and resulting sexually mature transgenic plants are self-pollinated to generate true breeding plants. The progeny from these plants become true breeding lines that are evaluated for, by way of example, increased biosynthesis of sulfur-containing compounds, preferably in the field, under a range of environmental conditions. The inventors contemplate that the present invention will find particular utility in the creation of transgenic plants of commercial interest including various grains, grasses, fibers, tubers, legumes, ornamental plants, cacti, succulents, fruits, berries, and vegetables, as well as a number of nut- and fruit-bearing trees and plants.

4.11 Recombinant Vectors Expressing MMT-Encoding Polynucleotides

One important embodiment of the invention is a recombinant vector that comprises a nucleic acid segment encoding one or more of the MMT-encoding genes disclosed herein. Such a vector may be transferred to and replicated in a prokaryotic or eukaryotic host, with bacterial cells being particularly preferred as prokaryotic hosts, and plant cells being particularly preferred as eukaryotic hosts.

In preferred embodiments, the recombinant vector comprises a nucleic acid segment encoding one or more of the MMT polypeptides disclosed herein. Highly preferred nucleic acid segments are those that comprise all, or substantially all of the coding regions that encode these polypeptides. Particularly preferred sequences are those that encode an MMT polypeptide that comprises an at least 7 contiguous amino acid sequence from any one of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8, and especially those sequences that encode an MMT

polypeptide that comprises all, or substantially all of the amino acid sequences disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.

Another important embodiment of the invention is a transformed host cell that expresses one or more of these recombinant vectors. The host cell may be either prokaryotic or eukaryotic, and particularly preferred host cells are those that express the nucleic acid segment(s) comprising the recombinant vector which encode one or more of the MMT polypeptides disclosed herein. Bacterial cells are particularly preferred as prokaryotic hosts, and plant cells are particularly preferred as eukaryotic hosts.

In another embodiment, the invention encompasses a method of using a nucleic acid segment that encodes one or more of the MMT polypeptides disclosed herein. The method generally comprises the steps of: (a) preparing a recombinant vector in which the gene is positioned under the control of a promoter; (b) introducing the recombinant vector into a host cell; (c) culturing the host cell under conditions effective to allow expression of the protein encoded by the gene; and (d) obtaining the expressed MMT peptide or polypeptide so produced.

A wide variety of ways are available for introducing a selected MMT-encoding polynucleotide into the microorganism host under conditions that allow for stable maintenance and expression of the gene. One can provide for DNA constructs that comprise (a) transcriptional and/or translational regulatory signals for expression of the gene, (b) the gene under their regulatory control, and (c) a DNA sequence homologous with a sequence in the host organism, whereby integration will occur, and/or a replication system which is functional in the host, whereby integration or stable maintenance will occur.

The transcriptional initiation signals will preferably include at least a first promoter and at least a first transcriptional initiation start site. In some instances, it may be desirable to provide for regulative expression of the polypeptide, where expression of the polypeptide will only occur after transformation into a suitable host cell, such as a transformed plant cell. This can be achieved with operators or a region binding to an activator or enhancers, which are capable of induction upon a change in the physical or chemical environment of the cell comprising the nucleic acid construct. For example, a temperature sensitive regulatory region may be employed, where the cells may be cultured in the laboratory without expression of MMT gene, but upon release into the environment, expression would begin.

Various manipulations may be employed for enhancing the expression of the messenger RNA, particularly by using an active promoter, as well as by employing sequences, which

enhance the stability of the messenger RNA. The transcriptional and translational termination region may preferably comprise one or more stop codon(s), terminator region(s), and optionally, one or more polyadenylation signal(s). A hydrophobic "leader" sequence may be employed at the amino terminus of the translated polypeptide sequence in order to promote secretion of the protein across the inner membrane.

In the direction of transcription, namely in the 5′ to 3′ direction of the coding or sense sequence, the construct will involve the transcriptional regulatory region, if any, and the promoter, where the regulatory region may be either 5′ or 3′ of the promoter, the ribosomal binding site, the initiation codon, the structural gene having an open reading frame in phase with the initiation codon, the stop codon(s), the polyadenylation signal sequence, if any, and the terminator region. This sequence as a double strand may be used by itself for transformation of a selected host cell, but will usually be included with a DNA sequence involving a marker, where the second DNA sequence may be joined to the MMT expression construct during introduction of the DNA into the host.

By a marker is intended a structural gene which provides for selection of those hosts which have been modified or transformed. The marker will normally provide for selective advantage, for example, providing for biocide resistance, *e.g.*, resistance to antibiotics or heavy metals; complementation, so as to provide prototropy to an auxotrophic host, or the like. Preferably, complementation is employed, so that the modified host may not only be selected, but may also be competitive in the field. One or more markers may be employed in the development of the constructs, as well as for modifying the host. The organisms may be further modified by providing for a competitive advantage against other wild-type microorganisms in the field. For example, genes expressing metal chelating agents, *e.g.*, siderophores may be introduced into the host along with the structural gene expressing the MMT polypeptide. In this manner, the enhanced expression of a siderophore may provide for a competitive advantage for the host that produces the MMT polypeptide, so that it may effectively compete with the wild-type microorganisms and stably occupy a niche in the environment.

Where no functional replication system is present, the construct will also include a sequence of at least 50 basepairs (bp), preferably at least about 100 bp, more preferably at least about 1000 bp, and usually not more than about 2000 bp of a sequence homologous with a sequence in the host. In this way, the probability of legitimate recombination is enhanced, so that the transgene will be integrated into the host DNA and stably maintained by the host.

Desirably, the transgene will be in close proximity to region of the host DNA where the integration is desired, thus providing for more efficient complementation as well permitting the stable integration of the transgene into the genome of the transformed host. Therefore, in the event that the transgene is lost, the resulting organism will be likely to also lose the complementing gene and/or the gene providing for the competitive advantage, so that it will be unable to compete in the environment with the gene retaining the intact construct.

A large number of transcriptional regulatory regions are available from a wide variety of microorganism hosts, such as bacteria, bacteriophage, cyanobacteria, algae, fungi, virus and the like. Various transcriptional regulatory regions include the regions associated with the *trp* gene, *lac* gene, *gal* gene, the λ_L and λ_R promoters, the *tac* promoter, the naturally-occurring promoters associated with MMT gene, where functional in the host. See for example, U.S. Patent Nos. 4,332,898; 4,342,832; and 4,356,270 (each of which is specifically incorporated herein by reference). The termination region may be the termination region normally associated with the transcriptional initiation region or a different transcriptional initiation region, so long as the two regions are compatible and functional in the host.

Where stable episomal maintenance or integration is desired, a plasmid will be employed which has a replication system that is functional in the selected host cell. The replication system may be derived from the chromosome, an episomal element normally present in the host or a different host, or a replication system from a virus that is stable in the host. A large number of standard cloning/expression plasmids are available, and their use to one of skill in the molecular biological arts in the preparation of transgenes and the like are well known. See for example, Olson *et al.* (1982); Bagdasarian *et al.* (1981), Baum *et al.*, 1990, and U.S. Patent Nos. 4,356,270; 4,362,817; 4,371,625, and 5,441,884, each incorporated specifically herein by reference.

The selected MMT gene can be introduced between the transcriptional and translational initiation region and the transcriptional and translational termination region, so as to be under the regulatory control of the initiation region. This construct will be included in a plasmid, which will include at least one replication system, but may include more than one, where one replication system is employed for cloning during the development of the plasmid and the second replication system is necessary for functioning in the ultimate host. In addition, one or more markers may be present, which have been described previously. Where integration is desired, the plasmid will desirably include a sequence homologous with the host genome.

The transformants can be isolated in accordance with conventional ways, usually employing a selection technique, which allows for selection of the desired organism as against unmodified organisms or transferring organisms, when present. The transformants then can be tested for activity. If desired, unwanted or ancillary DNA sequences may be selectively removed from the recombinant bacterium by employing site-specific recombination systems, such as those described in U.S. Patent No. 5,441,884 (specifically incorporated herein by reference).

In accordance with the present invention, nucleic acid sequences include and are not limited to DNA, including and not limited to cDNA and genomic DNA, genes; RNA, including and not limited to mRNA and tRNA; peptide nucleic acids (PNAs), ribozymes, antisense sequences, nucleosides, and suitable nucleic acid sequences such as those sequences encoding one or more of the MMT polypeptides disclosed herein.

As such the present invention also concerns DNA segments, that are free from total genomic DNA and that comprise one or more of the MMT-encoding polynucleotides disclosed herein. DNA segments encoding MMT polypeptide species may be obtained from native plant sources, or synthesized either partially or entirely *in vitro* using methods that are well known to those of skill in the art. Likewise, genes may be used that comprise all, or substantially all of a sequence that encodes an MMT polypeptide that retains its MMT enzyme activity in a suitably transformed host cell, when said cell is cultured under the appropriate conditions.

Included within the term "DNA segment", "nucleic acid segment" "polynucleotide" and "sequence region" are nucleic acid segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phagemids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified gene refers to a DNA segment which may include in addition to polypeptide encoding sequences, certain other elements such as, regulatory sequences, isolated substantially away from other naturally occurring genes or protein-encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein-, polypeptide- or peptide-encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, operon sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides or peptides.

"Isolated substantially away from other coding sequences" means that the gene of interest, in this case, a gene encoding a plant MMT polypeptide, forms the significant part of the

coding region of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or operon coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes, recombinant genes, synthetic linkers, or coding regions later added to the segment by the hand of man.

Particularly preferred are DNA sequences that encode one or more of the polypeptides disclosed in Table 9, and accordingly, sequences that have between about 70% and about 75% or between about 75% and about 80%, or more preferably between about 81% and about 90%, or even more preferably between about 91% or 92% or 93% and about 97% or 98% or 99% amino acid sequence identity or functional equivalence to the amino acid sequences disclosed in Table 9 will be highly desirable sequences for use in the practice of the present invention. It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5′ or 3′ sequences, and yet still be essentially as set forth in one of the sequences that encodes such an MMT polypeptide, so long as the sequence meets the criteria set forth above, including the maintenance of biological activity where expression of a functional polypeptide in a host cell is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5′ or 3′ portions of the coding region or may include various internal sequences, *i.e.*, introns, which are known to occur within genes.

The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

For certain application, relatively small contiguous nucleic acid sequences are preferable, such as those which are about 14 or 15 or 16 or 17 or 18 or 19, or 20, or 30-50, 51-80, 81-100 or so nucleotides in length. Alternatively, in some embodiments, and particularly those involving preparation of recombinant vectors, transformation of suitable host cells, and preparation of transgenic plant cell, longer nucleic acid segments are preferred, particularly those that include the entire coding region of one or more g=-polynucleotides encoding a functional MMT

polypeptide or an MMT-derived peptide sequence. As such, the preferred segments may include those that are up to about 20,000 or so nucleotides in length, or alternatively, shorter sequences such as those about 19,000, about 18,000, about 17,000, about 16,000, about 15,000, about 14,000, about 13,000, about 12,000, 11,000, about 10,000, about 9,000, about 8,000, about 7,000, about 6,000, about 5,000, about 4,500, about 4,000, about 3,500, about 3,000, about 2,500, about 2,000, about 1,500, about 1,000, about 500, or about 200 or so base pairs in length. Of course, these numbers are not intended to be exclusionary of all possible intermediate lengths in the range of from about 20,000 to about 15 nucleotides, as all of these intermediate lengths are also contemplated to be useful, and fall within the scope of the present invention. It will be readily understood that "intermediate lengths", in these contexts, means any length between the quoted ranges, such as 14, 15, 16, 17, 18, 19, 20, *etc.*; 21, 22, 23, 24, 25, 26, 27, 28, 29, *etc.*; 30, 31, 32, 33, 34, 35, 36, *etc.*; 40, 41, 42, 43, 44, *etc.*, 50, 51, 52, 53, *etc.*; 60, 61, 62, 63, *etc.*, 70, 80, 90, 100, 110, 120, 130, *etc.*; 200, 210, 220, 230, 240, 250, *etc.*; including all integers in the entire range from about 14 to about 10,000, including those integers in the ranges 200-500; 500-1,000; 1,000-2,000; 2,000-3,000; 3,000-5,000 and the like.

In one embodiment, a preferred polynucleotide will comprise a sequence of from about 1800 to about 18,000 base pair in length that encodes one or more native, mutagenized, or modified MMT polypeptides. The DNA segments of the present invention encompass biologically functional, equivalent peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally-equivalent proteins or peptides may be created *via* the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, *e.g.*, to introduce improvements to the activity of the protein, its expression, production, or persistence in a particular transformed host cell or to impart other desirable or beneficial characteristics to the mutagenized polypeptide.

If desired, one may also prepare fusion proteins and peptides, *e.g.*, where the peptide-coding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for purification or immunodetection purposes (*e.g.*, proteins that may be purified by affinity chromatography and enzyme label coding regions, respectively).

Recombinant vectors form further aspects of the present invention. Particularly useful vectors are contemplated to be those vectors in which the coding portion of the DNA segment, whether encoding a full-length protein, a substantially full-length or even a truncated or smaller polypeptide, is positioned under the control of at least a first promoter. The promoter may be in the form of the promoter that is naturally associated with a gene encoding peptides of the present invention, as may be obtained by isolating the 5′ non-coding sequences located upstream of the coding segment or exon, for example, using recombinant cloning and/or PCR™ technology, in connection with the compositions disclosed herein.

4.12 Transformed Plant Cells and Transgenic Plants Expressing MMT

In one embodiment, the invention provides a transgenic plant having incorporated into its genome a transgene that encodes an MMT polypeptide. A further aspect of the invention is a transgenic plant having incorporated into its genome a transgene that encodes such a polypeptide. Other embodiments of the invention also concern the progeny of such a transgenic plant, as well as its seed, the progeny from such seeds, and seeds arising from the second and subsequent generation plants derived from such a transgenic plant.

The invention also discloses and claims host cells, both native, and genetically engineered, which express one or more genes encoding all or substantially all of an MMT polypeptide to produce the encoded polypeptide(s) in a suitably transformed host cell, and in particular, in a transformed plant cell.

In yet another aspect, the present invention provides methods for producing a transgenic plant that expresses such a nucleic acid segment. The process of producing transgenic plants is well known in the art. In general, the method comprises transforming a suitable host cell with one or more DNA segments that contain a promoter operatively linked to a coding region that encodes one or more MMT polypeptides. Such a coding region is generally operatively linked to a transcription-terminating region, whereby the promoter is capable of driving the transcription of the coding region in the cell, and hence providing the cell the ability to produce the recombinant protein *in vivo*. Alternatively, in instances where it is desirable to control, regulate, or decrease the amount of a particular recombinant protein expressed in a particular transgenic cell, the invention also provides for the expression of an antisense oligonucleotide or other nucleic acid sequences that are complementary to the mRNA that encodes the expressed

polypeptide. The use of antisense mRNA as a means of controlling or decreasing the amount of a given protein of interest in a cell is well known in the art.

As used herein, the term "transgenic plant" is intended to refer to a plant that has incorporated DNA sequences, including but not limited to genes which are perhaps not normally present, DNA sequences not normally transcribed into RNA or translated into a protein ("expressed"), or any other genes or DNA sequences which one desires to introduce into the non-transformed plant, such as genes which may normally be present in the non-transformed plant but which one desires to either genetically engineer or to have altered expression.

It is contemplated that in some instances the genome of a transgenic plant of the present invention will have been augmented through the stable introduction of one or more transgenes, either native, synthetically modified, or mutated. In some instances, more than one transgene will be incorporated into the genome of the transformed host plant cell. Such is the case when more than one DNA segment is incorporated into the genome of such a plant. In certain situations, it may be desirable to have one, two, three, four, or even more MMT proteins (either native or recombinantly-engineered) incorporated and stably expressed in the transformed transgenic plant.

A preferred gene that may be introduced includes, for example, a DNA sequence from bacterial origin that encodes an MMT polypeptide, and particularly one or more of those described in SEQ ID NO:2, SEQ ID NO:4, and SEQ ID NO:6.

Means for transforming a plant cell and the preparation of a transgenic cell line are well known in the art, and are discussed herein. Vectors, plasmids, cosmids, bacterial artificial chromosomes (BACs), plant artificial chromosomes (PACs), yeast artificial chromosomes (YACs), and DNA segments for use in transforming such cells will, of course, generally comprise either the operons, genes, or gene-derived sequences of the present invention, either native, or synthetically-derived, and particularly those encoding the disclosed MMT polypeptides. These nucleic acid constructs can further include structures such as promoters, enhancers, polylinkers, or even gene sequences that have positively- or negatively-regulating activity upon the particular genes of interest as desired. The nucleic acid segment or gene may encode either a native or modified protein, which will be expressed in the resultant recombinant cells, and/or which will impart an improved phenotype to the regenerated plant

Such transgenic plants may be desirable for modulating the biosynthesis of sulfur-containing compounds in a population of monocotyledonous or dicotyledonous plants.

Particularly preferred plants include grains such as corn, wheat, rye, rice, barley, and oats; legumes such as beans, soybeans; tubers such as potatoes; fiber crops such as flax and cotton; turf and pasture grasses; ornamental plants; shrubs; trees; vegetables; berries; citrus crops, including oranges, tangerines, grapefruit, limes, lemons, and the like; fruits, cacti, succulents, and other commercially-important crops including greenhouse, garden and houseplants.

In a related aspect, the present invention also encompasses a seed produced by the transformed plant, a progeny from such seed, and a seed produced by the progeny of the original transgenic plant, produced in accordance with the above process. Such progeny and seeds will have one or more MMT-encoding transgene(s) stably incorporated into its genome, and such progeny plants will inherit the traits afforded by the introduction of a stable transgene in Mendelian fashion. All such transgenic plants having incorporated into their genome transgenic DNA segments encoding one or more MMT polypeptides are aspects of this invention.

4.13 Isolating Homologous Gene and Gene Fragments Encoding MMT

The polynucleotide sequences of the subject invention include not only full-length sequences but also fragments of these sequences, (including *e.g.,* fusion proteins), which retain the MMT enzymatic activity of the sequences specifically exemplified herein in SEQ ID NO:2, SEQ ID NO:4, and SEQ ID NO:6.

It should be apparent to a person skilled in this art that the various genetic constructs encoding MMT polypeptides can be identified and obtained through several means. The MMT-encoding genes, or portions thereof, may be obtained from a culture depository, or constructed synthetically, for example, by use of a gene machine. Variations of these genes may be readily constructed using standard techniques for making point mutations. Also, fragments of these genes can be made using commercially available exonucleases or endonucleases according to standard procedures. For example, enzymes such as *Bal*31 or site-directed mutagenesis can be used to systematically cut off nucleotides from the ends of these genes. Also, genes or gene fragments that encode biologically active polypeptides may be obtained using a variety of other restriction enzymes. Proteases may be used to directly obtain active fragments of these constructs.

Equivalent polypeptides and/or polynucleotides encoding these equivalent polypeptides can also be isolated from DNA libraries using the teachings provided herein. For example, antibodies to the polypeptides disclosed and claimed herein can be used to identify and isolate

other similar or related polypeptides from a mixture of proteins. These antibodies can then be used to specifically identify equivalent polypeptides possessing the desired characteristics by a variety of methodologies including, *e.g.*, immunoprecipitation, enzyme linked immunoassay (ELISA), and/or Western blotting.

A further method for identifying the polypeptides and polynucleotides of the subject invention is through the use of oligonucleotide probes. These probes are nucleotide sequences having a detectable label. As is well known in the art, if the probe molecule and nucleic acid sample hybridize by forming a strong bond between the two molecules, it can be reasonably assumed that the probe and sample are essentially identical. The probe's detectable label provides a means for determining in a known manner whether hybridization has occurred. Such a probe analysis provides a rapid method for identifying genes of the subject invention.

The nucleotide segments that are used as probes according to the invention may be synthesized by use of nucleic acid synthesizers using standard procedures. In the use of the nucleotide segments as probes, the particular probe is labeled with any suitable label known to those skilled in the art, including radioactive and non-radioactive labels. Typical radioactive labels include ^{32}P, ^{125}I, ^{35}S, or the like. A probe labeled with a radioactive isotope can be constructed from a nucleotide sequence complementary to the DNA sample by a conventional nick translation reaction, using a DNase and DNA polymerase. The probe and sample can then be combined in a hybridization buffer solution and held at an appropriate temperature until annealing occurs. Thereafter, the membrane is washed free of extraneous materials, leaving the sample and bound probe molecules typically detected and quantified by autoradiography and/or liquid scintillation counting.

Non-radioactive labels include, for example, ligands such as biotin or thyroxine, as well as enzymes such as hydrolases or peroxidases, or the various chemiluminescers such as luciferin, or fluorescent compounds like fluorescein and its derivatives. The probe may also be labeled at both ends with different types of labels for ease of separation, as, for example, by using an isotopic label at the end mentioned above and a biotin label at the other end.

Duplex formation and stability depend on substantial complementarity between the two strands of a hybrid, and, as noted above, a certain degree of mismatch can be tolerated. Therefore, the probes of the subject invention include mutations (both single and multiple), deletions, insertions of the described sequences, and combinations thereof, wherein said mutations, insertions and deletions permit formation of stable hybrids with the target

polynucleotide of interest. Mutations, insertions, and deletions can be produced in a given polynucleotide sequence in many ways, by methods currently known to an ordinarily skilled artisan, and perhaps by other methods which may become known in the future.

The potential variations in the probes listed are due, in part, to the redundancy of the genetic code. Because of the redundancy of the genetic code, *i.e.* more than one coding nucleotide triplet (codon) can be used for most of the amino acids used to make proteins. Therefore different nucleotide sequences can code for a particular amino acid. Thus, the amino acid sequences of the disclosed polypeptides can be prepared by equivalent nucleotide sequences encoding the same amino acid sequence of the protein or peptide. Accordingly, the subject invention includes such equivalent nucleotide sequences. Also, inverse or complement sequences are an aspect of the subject invention and can be readily used by a person skilled in this art. In addition it has been shown that proteins of identified structure and function may be constructed by changing the amino acid sequence if such changes do not alter the protein secondary structure (Kaiser and Kezdy, 1984). Thus, the subject invention includes mutants of the amino acid sequence depicted herein that do not alter the protein secondary structure, or if the structure is altered, the biological activity is substantially retained. Further, the invention also includes mutants of organisms hosting all or part of one or more of the DNA constructs of the invention. Such mutants can be made by techniques well known to persons skilled in the art. For example, UV irradiation can be used to prepare mutants of host organisms. Likewise, such mutants may include asporogenous host cells that also can be prepared by procedures well known in the art.

4.14 Peptide Nucleic Acid Compositions

In certain embodiments, the inventors contemplate the use of peptide nucleic acids (PNAs) in the practice of the methods of the invention. PNAs are DNA analogs that mimic the structure of the polynucleotide, in which the nucleobases are attached to a pseudopeptide backbone (Good and Nielsen, 1997). PNAs can be utilized in a number of methods that traditionally have used RNAs or DNAs (U.S. Patent No. 5,786,461; U.S. Patent No. 5,773,571, U.S. Patent No. 5,766,855; U.S. Patent No. 5,736,336; U.S. Patent No. 5,719,262; and U.S. Patent No. 5,539,082, each specifically incorporated herein by reference in its entirety). Often PNA sequences perform better in techniques than the corresponding RNA or DNA sequences

and have utilities that are not inherent to RNA or DNA. Methods of making, and using PNAs are also found in Corey (1997).

PNAs when delivered within cells have the potential to be general sequence-specific regulators of gene expression. Reviews of PNAs and their use as antisense and anti-gene agents exist (Nielsen *et al.*, 1993b; Hanvey *et al.*, 1992; and Good and Nielsen, 1997). Other applications of PNAs include use in DNA strand invasion (Nielsen *et al.*, 1991), antisense inhibition (Hanvey *et al.*, 1992), mutational analysis (Orum *et al.*, 1993), enhancers of transcription (Mollegaard *et al.*, 1994), nucleic acid purification (Orum *et al.*, 1995), isolation of transcriptionally active genes (Boffa *et al.*, 1995), blocking of transcription factor binding (Vickers *et al.*, 1995), genome cleavage (Veselkov *et al.*, 1996), biosensors (Wang *et al.*, 1996), *in situ* hybridization (Thisted *et al.*, 1996), and in an alternative to Southern blotting (Perry-O'Keefe, 1996).

4.15 MMT-Specific Antibody Compositions and Formulations Thereof

Means for preparing and characterizing antibodies are well known in the art (see, *e.g.*, Harlow and Lane, 1988; incorporated herein by reference). The methods for generating monoclonal antibodies (mAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic composition in accordance with the present invention and collecting antisera from that immunized animal. A wide range of animal species can be used for the production of antisera. Typically the animal used for production of anti-antisera is a rabbit, a mouse, a rat, a hamster, a guinea pig or a goat. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin can also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, *m*-maleimidobencoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

As is also well known in the art, the immunogenicity of a particular immunogen composition can be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Exemplary and preferred adjuvants include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis)*, incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

The amount of immunogen composition used in the production of polyclonal antibodies varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points following immunization. A second, booster, injection may also be given. The process of boosting and titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the immunized animal can be bled and the serum isolated and stored, and/or the animal can be used to generate mAbs.

mAbs that are specific for an MMT peptide epitope, peptide, or polyepeptide may be readily prepared through use of well-known techniques, such as those exemplified in U. S. Patent 4,196,265, incorporated herein by reference. Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, *e.g.*, a purified or partially purified protein, polypeptide or peptide. The immunizing composition is administered in a manner effective to stimulate antibody producing cells. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep frog cells is also possible. The use of rats may provide certain advantages (Goding, 1986), but mice are preferred, with the BALB/c mouse being most preferred as this is most routinely used and generally gives a higher percentage of stable fusions.

Following immunization, somatic cells with the potential for producing antibodies, specifically B-lymphocytes (B-cells), are selected for use in the mAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible. Often, a panel of animals will have been immunized and the spleen of animal with the highest antibody titer removed and the spleen

lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately about 5×10^7 to about 2×10^8 lymphocytes.

The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render then incapable of growing in certain selective media which support the growth of only the desired fused cells (hybridomas).

Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding, 1986; Campbell, 1984). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8, X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used is the 8-azaguanine-resistant mouse murine myeloma SP2/0 non-producer cell line.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 ratio, though the ratio may vary from about 20:1 to about 1:1, respectively, in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described (Kohler and Milstein, 1975; 1976), and those using polyethylene glycol (PEG), such as 37% (vol./vol.) PEG, by Gefter *et al.* (1977). The use of electrically induced fusion methods is also appropriate (Goding, 1986).

Fusion procedures usually produce viable hybrids at low frequencies, about 1×10^{-6} to about 1×10^{-8}. However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is generally one that contains an agent that blocks the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block *de novo* synthesis of both purines and

pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key enzymes of the salvage pathway, *e.g.*, hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B-cells can operate this pathway, but they have a limited life span in culture and generally die within about two wk. Therefore, the only cells that can survive in the selective media are those hybrids formed from myeloma and B-cells.

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three wk) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

The selected hybridomas would then be serially diluted and cloned into individual antibody-producing cell lines, which clones can then be propagated indefinitely to provide mAbs. The cell lines may be exploited for mAb production in two basic ways. A sample of the hybridoma can be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion. The injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, can then be tapped to provide mAbs in high concentration. The individual cell lines may also be cultured *in vitro*, where the mAbs are naturally secreted into the culture medium from which they can be readily obtained in high concentrations. mAbs produced by either means may be further purified, if desired, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography.

4.16 EPITOPIC CORE SEQUENCES

The present invention is also directed to MMT polypeptide compositions, free from total cells and other polypeptides, which comprise a purified MMT polypeptide which incorporates an

epitope that is immunologically cross-reactive with one or more of the MMT-specific antibodies of the present invention.

As used herein, the term "incorporating an epitope(s) that is immunologically cross-reactive with one or more anti-MMT antibodies" is intended to refer to a peptide or protein antigen which includes a primary, secondary or tertiary structure similar to an epitope located within an MMT polypeptide. The level of similarity will generally be to such a degree that monoclonal or polyclonal antibodies directed against the MMT polypeptide will also bind to, react with, or otherwise recognize, the cross-reactive peptide or protein antigen. Various immunoassay methods may be employed in conjunction with such antibodies, such as, for example, Western blotting, ELISA, RIA, and the like, all of which are known to those of skill in the art.

The identification of MMT epitopes and/or their functional equivalents, suitable for use in vaccines is a relatively straightforward matter. For example, one may employ the methods of Hopp, as taught in U.S. Patent 4,554,101, incorporated herein by reference, which teaches the identification and preparation of epitopes from amino acid sequences on the basis of hydrophilicity. The methods described in several other papers, and software programs based thereon, can also be used to identify epitopic core sequences (see, for example, Jameson and Wolf, 1988; Wolf *et al.*, 1988; U. S. Patent 4,554,101). The amino acid sequence of these "epitopic core sequences" may then be readily incorporated into peptides, either through the application of peptide synthesis or recombinant technology.

Preferred peptides for use in accordance with the present invention will generally be on the order of about 5 to about 25 amino acids in length, and more preferably about 8 to about 20 amino acids in length. It is proposed that shorter antigenic peptide sequences will provide advantages in certain circumstances, for example, in the preparation of vaccines or in immunologic detection assays. Exemplary advantages include the ease of preparation and purification, the relatively low cost and improved reproducibility of production, and advantageous biodistribution.

It is proposed that particular advantages of the present invention may be realized through the preparation of synthetic peptides which include modified and/or extended epitopic/immunogenic core sequences which result in a "universal" epitopic peptide directed to MMT-related sequences. It is proposed that these regions represent those which are most likely

to promote T-cell or B-cell stimulation in an animal, and, hence, elicit specific antibody production in such an animal.

An epitopic core sequence, as used herein, is a relatively short stretch of amino acids that is "complementary" to, and therefore will bind, antigen binding sites on MMT epitope-specific antibodies. Additionally or alternatively, an epitopic core sequence is one that will elicit antibodies that are cross-reactive with antibodies directed against the peptide compositions of the present invention. It will be understood that in the context of the present disclosure, the term "complementary" refers to amino acids or peptides that exhibit an attractive force towards each other. Thus, certain epitope core sequences of the present invention may be operationally defined in terms of their ability to compete with or perhaps displace the binding of the desired protein antigen with the corresponding protein-directed antisera.

In general, the size of the polypeptide or peptide antigen is not believed to be particularly crucial, so long as it is at least large enough to carry the identified core sequence or sequences. The smallest useful core sequence expected by the present disclosure would generally be on the order of about 5 amino acids in length, with sequences on the order of 8 or 25 being more preferred. Thus, this size will generally correspond to the smallest peptide antigens prepared in accordance with the invention. However, the size of the antigen may be larger where desired, so long as it contains a basic epitopic core sequence.

The identification of epitopic core sequences is known to those of skill in the art, for example, as described in U. S. Patent 4,554,101, incorporated herein by reference, which teaches the identification and preparation of epitopes from amino acid sequences on the basis of hydrophilicity. Moreover, numerous computer programs are available for use in predicting antigenic portions of proteins (see *e.g.*, Jameson and Wolf, 1988; Wolf *et al.*, 1988). Computerized peptide sequence analysis programs (*e.g.*, DNAStar™ software, DNAStar, Inc., Madison, WI) may also be useful in designing synthetic epitopes and epitope analogs in accordance with the present disclosure.

In certain embodiments, particular advantages may be realized through the preparation of synthetic MMT peptides that include epitopic/immunogenic core sequences. These epitopic core sequences may be identified as hydrophilic and/or mobile regions of the polypeptides or those that include a T cell motif. It is known in the art that such regions represent those that are most likely to promote B cell or T cell stimulation, and, hence, elicit specific antibody production.

To confirm that a polypeptide is immunologically cross-reactive with, or a biological functional equivalent of, one or more epitopes of the disclosed MMT peptides is also a straightforward matter. This can be readily determined using specific assays, *e.g.*, of a single proposed epitopic sequence, or using more general screens, *e.g.*, of a pool of randomly generated synthetic peptides or protein fragments. The screening assays may be employed to identify either equivalent antigens or cross-reactive antibodies. In any event, the principle is the same, *i.e.* based upon competition for binding sites between antibodies and antigens.

Suitable competition assays that may be employed include protocols based upon immunohistochemical assays, ELISAs, RIAs, Western or dot blotting and the like. In any of the competitive assays, one of the binding components, generally the known element, such as an MMT or MMT-derived peptide, or a known antibody, will be labeled with a detectable label and the test components, that generally remain unlabeled, will be tested for their ability to reduce the amount of label that is bound to the corresponding reactive antibody or antigen.

As an exemplary embodiment, to conduct a competition study between MMT and any test antigen, one would first label MMT with a detectable label, such as, *e.g.*, biotin or an enzymatic, radioactive or fluorogenic label, to enable subsequent identification. One would then incubate the labeled antigen with the other, test, antigen to be examined at various ratios (*e.g.*, 1:1, 1:10 and 1:100) and, after mixing, one would then add the mixture to a known antibody. Preferably, the known antibody would be immobilized, *e.g.*, by attaching to an ELISA plate. The ability of the mixture to bind to the antibody would be determined by detecting the presence of the specifically bound label. This value would then be compared to a control value in which no potentially competing (test) antigen was included in the incubation.

The assay may be any one of a range of immunological assays based upon hybridization, and the reactive antigens would be detected by means of detecting their label, *e.g.*, using streptavidin in the case of biotinylated antigens or by using a chromogenic substrate in connection with an enzymatic label or by simply detecting a radioactive or fluorescent label.

The reactivity of the labeled antigen, *e.g.*, an MMT-derived peptide, in the absence of any test antigen would be the control high value. The control low value would be obtained by incubating the labeled antigen with an excess of unlabeled antigen, when competition would occur and reduce binding. A significant reduction in labeled antigen reactivity in the presence of a test antigen is indicative of a test antigen that is "cross-reactive", *i.e.* that has binding affinity

for the same antibody. "A significant reduction", in terms of the present application, may be defined as a reproducible (*i.e.* consistently observed) reduction in binding.

In addition to the peptidyl compounds described herein, the inventors also contemplate that other sterically similar compounds may be formulated to mimic the key portions of the peptide structure. Such compounds, which may be termed peptidomimetics, may be used in the same manner as the peptides of the invention and hence are also functional equivalents. The generation of a structural functional equivalent may be achieved by the techniques of modeling and chemical design known to those of skill in the art. It will be understood that all such sterically similar constructs fall within the scope of the present invention.

Syntheses of epitopic sequences, or peptides which include an antigenic epitope within their sequence, are readily achieved using conventional synthetic techniques such as the solid phase method (*e.g.*, through the use of a commercially-available peptide synthesizer such as an Applied Biosystems Model 430A Peptide Synthesizer). Peptide antigens synthesized in this manner may then be aliquoted in predetermined amounts and stored in conventional manners, such as in aqueous solutions or, even more preferably, in a powder or lyophilized state pending use.

In general, due to the relative stability of peptides, they may be readily stored in aqueous solutions for fairly long periods of time if desired, *e.g.*, up to six months or more, in virtually any aqueous solution without appreciable degradation or loss of antigenic activity. However, where extended aqueous storage is contemplated it will generally be desirable to include agents including buffers such as Tris or phosphate buffers to maintain a pH of about 7.0 to about 7.5. Moreover, it may be desirable to include agents that will inhibit microbial growth, such as sodium azide or Merthiolate. For extended storage in an aqueous state it will be desirable to store the solutions at 4°C, or more preferably, frozen. Of course, where the peptides are stored in a lyophilized or powdered state, they may be stored virtually indefinitely, *e.g.*, in metered aliquots that may be rehydrated with a predetermined amount of water (preferably distilled) or buffer prior to use.

4.17 Biological Functional Equivalents

Modification and changes may be made in the structure of the MMT-specific genes, promoters, genetic constructs, plasmids, and/or polypeptides of the present invention and still obtain functional molecules that possess the desirable biologically-active characteristics. The

following is a discussion based upon changing the amino acids of a protein to create an equivalent, or even an improved, second-generation molecule. In particular embodiments of the invention, mutated polynucleotides and/or polypeptides are contemplated to be useful for increasing the biosynthesis of sulfur-containing compounds in a transformed cell, and consequently increasing the activity and/or expression of the recombinant MMT-encoding transgene in a plant cell. The amino acid changes may be achieved by changing the codons of the DNA sequence, according to the codons given in Table 6.

TABLE 6

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic Acid	Asp	D	GAC	GAU				
Glutamic Acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences that encode said peptides without appreciable loss of their biological utility or activity.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporate herein by reference). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics (Kyte and Doolittle, 1982), these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (–0.4); threonine (–0.7); serine (–0.8); tryptophan (–0.9); tyrosine (–1.3); proline (–1.6); histidine (–3.2); glutamate (–3.5); glutamine (–3.5); aspartate (–3.5); asparagine (–3.5); lysine (–3.9); and arginine (–4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.*, still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent No. 4,554,101, specifically incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed

by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U.S. Patent No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (–0.4); proline (–0.5 ± 1); alanine (–0.5); histidine (–0.5); cysteine (–1.0); methionine (–1.3); valine (–1.5); leucine (–1.8); isoleucine (–1.8); tyrosine (–2.3); phenylalanine (–2.5); tryptophan (–3.4).

It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions which take several of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

4.18 Antisense Oligonucleotides Targeted to mRNA

In certain embodiments, the inventors contemplate the use of antisense compositions to negatively regulate the expression of an MMT-encoding gene sequence in a host cell. The end result of the flow of genetic information is the synthesis of protein. DNA is transcribed by polymerases into messenger RNA and translated on the ribosome to yield a folded, functional protein. Thus, even from this simplistic description of an extremely complex set of reactions, it is obvious that there are several steps along the route where protein synthesis can be inhibited. The native DNA segment encoding a MMT, as all such mammalian DNA strands, has two strands: a sense strand and an antisense strand held together by hydrogen bonding. The messenger RNA encoding a MMT has the same nucleotide sequence as the sense DNA strand except that the DNA thymidine is replaced by uridine. Thus, antisense nucleotide sequences will bind to the mRNA encoding an MMT polypeptide and inhibit production of the protein.

The targeting of antisense oligonucleotides to bind mRNA is one mechanism to shut down protein synthesis. For example, the synthesis of polygalactauronase and the muscarine type-2 acetylcholine receptor are inhibited by antisense oligonucleotides directed to their respective mRNA sequences (U. S. Patent 5,739,119 and U. S. Patent 5,759,829, U. S. Patent 5,801,154; U. S. Patent 5,789,573; U. S. Patent 5,718,709 and U. S. Patent 5,610,288, each specifically incorporated herein by reference in its entirety).

In illustrative embodiments, antisense oligonucleotides may be prepared which are complementary nucleic acid sequences that can recognize and bind to target genes or the transcribed mRNA, resulting in the arrest and/or inhibition of deoxyribonucleic acid (DNA) transcription or translation of the messenger ribonucleic acid (mRNA). These oligonucleotides can be expressed within a host cell that normally expresses MMT-specific mRNA to reduce or inhibit the expression of this mRNA. Thus, the oligonucleotides may be useful for reducing the level of MMT polypeptide in a suitably transformed host cell or transgenic plant.

The native nucleic acid segment encoding MMT has, as do all such plant DNAs, two strands: a sense strand and an antisense strand held together in a duplex formation by hydrogen bonding. The messenger RNA (mRNA) encoding MMT has the same nucleotide sequence as the sense DNA strand except that the thymidine in DNA is replaced by uridine in DNA. Thus, preferred antisense oligonucleotide compositions for use in the practice of the present invention are those sequences that specifically bind to the mRNA coding for MMT and that inhibit or reduce the expression of the MMT polypeptide encoding by that mRNA.

The present invention provides a composition comprising at least a first oligonucleotide of at least about 9 to about 45 or so bases in length, wherein the oligonucleotide specifically binds to a portion of mRNA expressed from a gene encoding a plant MMT polypeptide, and further wherein binding of the oligonucleotide to the mRNA is effective in decreasing the activity of or reducing the quantity of the MMT enzyme in a host plant cell expressing the mRNA.

In certain aspects of the invention, the oligonucleotide comprises deoxyribonucleic acid, ribonucleic acid, or peptide-nucleic acid. In particular embodiments, the oligonucleotide comprises a sequence of at least nine, at least ten, at least eleven, at least twelve, at least thirteen, or at least fourteen, up to and including the full-length contiguous sequences from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID

NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:25. When longer antisense molecules are required, one may employ an oligonucleotide that comprises a sequence of at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, or at least twenty, up to and including the full-length contiguous sequences from SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7. Such antisense molecules may comprise even longer contiguous nucleotide sequences, such as those comprising about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, or about 30 or so contiguous nucleotides from any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:25.

4.19 DEFINITIONS

In accordance with the present invention, nucleic acid sequences include and are not limited to DNA (including and not limited to genomic or extragenomic DNA), genes, RNA (including and not limited to mRNA and tRNA), nucleosides, and suitable nucleic acid segments either obtained from native sources, chemically synthesized, modified, or otherwise prepared by the hand of man. The following words and phrases have the meanings set forth below.

A, an: In accordance with long standing patent law convention, the words "a" and "an" when used in this application, including the claims, denotes "one or more".

Expression: The combination of intracellular processes, including transcription and translation undergone by a coding DNA molecule such as a structural gene to produce a polypeptide.

Promoter: A recognition site on a DNA sequence or group of DNA sequences that provide an expression control element for a structural gene and to which RNA polymerase specifically binds and initiates RNA synthesis (transcription) of that gene.

Regeneration: The process of growing a plant from a plant cell (*e.g.*, plant protoplast or explant).

Structural gene: A gene that is expressed to produce a polypeptide.

Transformation: A process of introducing an exogenous DNA sequence (*e.g.*, a vector, a recombinant DNA molecule) into a cell or protoplast in which that exogenous DNA is incorporated into a chromosome or is capable of autonomous replication.

Transformed cell: A cell whose DNA has been altered by the introduction of an exogenous DNA molecule into that cell.

Transgenic cell: Any cell derived or regenerated from a transformed cell or derived from a transgenic cell. Exemplary transgenic cells include plant calli derived from a transformed plant cell and particular cells such as leaf, root, stem, *e.g.*, somatic cells, or reproductive (germ) cells obtained from a transgenic plant.

Transgenic plant: A plant or progeny thereof derived from a transformed plant cell or protoplast, wherein the plant DNA contains an introduced exogenous DNA molecule not originally present in a native, non-transgenic plant of the same strain. The terms "transgenic plant" and "transformed plant" have sometimes been used in the art as synonymous terms to define a plant whose DNA contains an exogenous DNA molecule. However, it is thought more scientifically correct to refer to a regenerated plant or callus obtained from a transformed plant cell or protoplast as being a transgenic plant, and that usage will be followed herein.

Vector: A DNA molecule capable of replication in a host cell and/or to which another DNA segment can be operatively linked so as to bring about replication of the attached segment. A plasmid is an exemplary vector.

5.0 EXAMPLES

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

5.1 EXAMPLE 1

All flowering plants produce S-methylmethionine (SMM) from methionine (Met) and have a separate mechanism to convert SMM back to Met. The functions of SMM and the

reasons for its interconversion with Met are not known. This example demonstrates that L-SMM is a major constituent of the phloem sap. L-SMM was likewise a prominent labeled product in phloem exudates obtained by EDTA treatment of detached leaves of *Poaceae*, *Fabaceae*, *Asteraceae*, *Brassicaceae*, and *Cucurbitaceae* that were given L-[^{35}S] Met.

cDNA clones for the enzyme that catalyzes SMM synthesis (S-adenosylMet:Met S-methyltransferase) were isolated from *Wollastonia biflora*, maize, and *Arabidopsis*. The deduced amino acid sequences revealed the expected methyltransferase domain (around 30 residues, sited at the N-terminus), plus an 800-residue C-terminal region sharing significant similarity with pyridoxal-5'-phosphate (PLP)-dependent enzymes. These data indicated that SMM has a previously unrecognized but often-major role in sulfur transport in flowering plants and that evolution of SMM synthesis in this group involved a gene fusion event. The resulting bipartite enzyme is unlike any other methyltransferase.

Because SMM is a major and common phloem constituent, the cDNAs for MMT, the SMM-synthesizing enzyme, were isolated and characterized from three diverse flowering plants. The deduced MMT amino acid sequences define a novel type of methyltransferase.

5.1.1 Experimental Methods

5.1.1.1 Plant Materials

W. biflora (L.) DC genotype H was grown as described (Trossat *et al.*, 1996). Other plants (for exudate collection by the EDTA method) were grown in a growth chamber (16-hour day, PPFD 200-300 μmol m^{-2} s^{-1}, 22°C day/18°C night). These plants were *Arabidopsis thaliana* (L.) Heynh. ecotype RLD, broad bean (*Vicia faba* L.), canola (*Brassica napus* L.), cucumber (*Cucumis sativus* L.), white lupine (*Lupinus albus* L.), maize (*Zea mays* L.), radish (*Raphanus sativus* L.), soybean (*Glycine max* L.), wheat (*T. aestivum* L. cv Bob White, and zucchini (*Cucurbita pepo* L.); all were used for studies when at the flowering or seed-filling stage.

5.1.1.2 Collection of Phloem Exudates

Labeled phloem sap was obtained by EDTA-enhanced exudation (King *et al.*, 1974). A 5-μl droplet containing [^{35}S]Met (20μCi) was applied to beveled tips of attached leaves, and the plants were left in the growth chamber for 2 h. The labeled leaf was then severed, placed with its base in 1 ml of 5 mM NA_2EDTA, pH 7.0, and held in a humid chamber in darkness for 20 h. to permit exudation.

5.1.1.3 Analysis of ^{35}S-Labeled Compounds

Leaves were extracted as described previously. Leaf extracts and phloem exudates were fractionated by passage through 1 ml columns of Dowex-1 (OH^-) and BioRex-70 (H^+) arranged in series; these respectively retain neutral + acidic amino acids, and SMM (James *et al.*, 1995a; 1995b). Subsequent separations were by TLC system 1 and thin layer electrophoresis system 2 (James *et al.*, 1995a; 1995b), or by TLC on cellulose plates developed with *n*-propanol:formic acid:water (20:1:5, vol./vol./vol.). Radioactivity was detected by autoradiography, and amino acids with ninhydrin. The identity of [^{35}S]SMM was established by co-migration with authentic SMM in the above systems; by decomposition upon treating with 1 M NaOH at 100°C for 2 h (White, 1982); and by conversion to the α-hydroxy acid with nitrous acid (James *et al.*, 1995a; 1995b). The complete methods for isolation, purification, and assay of MMT enzymatic activity is given in James *et al.*, (1995a, 1995b) and Pimenta *et al.*, (1998).

5.1.1.4 cDNA Cloning and Sequence Analysis

Peptides were obtained from *W. biflora* MMT using endoproteinase Lys C or trypsin, separated by HPLC, and sequenced by Edman degradation or by MS. mRNA was isolated from *W. biflora* leaves as described (Rathinasabapathi *et al.*, 1997) and used to construct a cDNA library in the Uni- Zap XR vector (Stratagene).

The degenerate PCR™ primers 5'-AARTTYTTRAAYGCIAAYATHATG-3' (SEQ ID NO:54) and 5'-TTRAAICCIACYTCRGCYTC-3' (SEQ ID NO:55) (corresponding to the ends of the peptide KFLNANIMSIPTEAEVGFK (SEQ ID NO:56), were used to amplify a 56-bp DNA fragment using the cDNA library (2×10^8 pfu per 50-µl reaction) as template. Sequencing confirmed that this fragment encoded the expected residues; a specific primer: 5'-ATCATGTCTATCCCTACAGA-3' (SEQ ID NO:57) from its center was then used with the vector T7 primer to amplify the 1.1-kb 3'-region of the MMT cDNA. The 1.1-kb fragment was then used to screen the library (3×10^4 plaques); this yielded greater than 100 MMT cDNAs, all truncated, of which the longest was 3.1-kb in length. The 5'-terminus was obtained by RACE, using the GIBCO-BRL kit. Cloned RACE products from four independent reactions were sequenced; a clone with no errors (judged from the sequence consensus) was used to construct the complete cDNA by fusion at the *Nhe*I site.

Arabidopsis EST 205D23T7 (GenBank Accession No. H77211) was obtained from the *Arabidopsis* Biological Resource Center and shown by sequencing to encode the 1.2-kb 3'-region of a protein homologous to *W. biflora* MMT. The 1.2-kb cDNA was then used to isolate a full-length cDNA from an *Arabidopsis* leaf library in Lambda Uni-Zap XR. A maize EST was also recovered from a maize genetic library.

Both strands of cDNAs were sequenced using the ABI PRISM dye terminator cycle sequencing Ready Reaction and an automated sequencer (ABI Model 373). Sequence alignments were made using Multalin (Corpet, 1988) or Clustal W v1.7 (Thompson *et al.*, 1994) programs and shaded with BOXSHADE. Homology searches were made using the BLAST v2.0.9 program according to the method previously described (Altschul *et al.*, 1997).

5.1.2 RESULTS

5.1.2.1 LEAVES OF DIVERSE FLOWERING PLANTS EXPORT SMM IN PHLOEM

To find whether SMM is a constituent of phloem sap in other species, EDTA was used to enhance phloem exudation from cut leaf bases or petioles. EDTA chelates the Ca^{2+} required for callose formation and thereby blocks the scaling of cut sieve tubes (King *et al.*, 1974). The sap obtained by this technique is comparable in composition to that from severed stylets (Weibull *et al.*, 1990; Valle *et al.*, 1998). A tracer dose of [^{35}S]Met was applied to the tips of leaves from species representing five diverse families, and the corresponding phloem exudates were analyzed for [^{35}S]SMM (FIG. 2A). Amino acid exudation was measured from matching unlabeled leaves, plus or minus EDTA, as a check on the technique. EDTA enhanced amino acid exudation by an average of 9-fold (FIG. 2B), consistent with the exudates coming mainly from cut sieve tubes (King *et al.*, 1974). [^{35}S]SMM was detected in the exudates from all eleven species tested, and in five of them accounted for >35% of the total label (FIG. 2A). In these studies, the applied tracer [^{35}S]Met may have entered the phloem directly whereas ^{35}S label could only have reached SMM after isotope dilution by endogenous Met and SMM pools. These pools are far larger than the applied [^{35}S]Met doses and vary in size with species and leaf age. Thus, the low proportion of [^{35}S]SMM in the exudates of some species by no means necessarily connotes low chemical levels of SMM.

5.1.2.2 cDNA Cloning of MMTs from Other Species

The widespread occurrence of SMM in phloem sap led to the isolation of cDNAs for the SMM-synthesizing enzyme, MMT, from a variety of plant species including *Wollastonia biflora*. *W. biflora* MMT was digested to obtain peptides, 14 of which were sequenced. Degenerate polymerase chain reaction (PCR™) primers matching the ends of a 19-residue peptide (see methods) were used to amplify the corresponding 56 bp DNA sequence, with a *W. biflora* cDNA library as template. A primer specific for the central part of this 56-bp sequence, together with one from the vector, enabled PCR™ amplification of a 1.1-kb fragment comprising the 3'-terminal region of MMT. This fragment was used to screen the library, which led to isolation of a 3.1-kb MMT cDNA that lacked the 5' region. The missing region was obtained by rapid amplification of cDNA ends (RACE). The complete cDNA (SEQ ID NO:1) encodes a 1088-residue protein (SEQ ID NO:2) having a calculated mass of 121.6 kDa (FIG. 3A, FIG. 3B and FIG. 3C). The deduced amino acid sequence includes all the MMT peptide fragmentss that were sequenced, establishing its authenticity. No N-terminal signal sequence was recognizable, consistent with the exclusively cytosolic localization of *W. biflora* MMT (Trossat *et al.*, 1996).

Searches of expressed sequence tag (EST) databases using the *W. biflora* nucleotide and amino acid sequences revealed matches with greater than ten maize and two *Arabidopsis* ESTs.

One *Arabidopsis* EST (1.2-kb) was sequenced, confirmed to encode a polypeptide similar to *W. biflora* MMT, and used to screen an *Arabidopsis* cDNA library. This identified a clone with a 3.4-kb insert containing a gene (SEQ ID NO:3) that encoded a 1071-residue polypeptide (SEQ ID NO:4) that had a 67% amino acid identity (81% similarity) to the *W. biflora* MMT polypeptide (FIG. 3A, FIG. 3B and FIG. 3C).

The longest maize EST (3.4-kb) was sequenced and the resulting maize gene (SEQ ID NO:5) was found to encode a 1091-residue polypeptide (SEQ ID NO:6) with 62% amino acid identity (77% similarity) to *W. biflora* MMT (FIG. 3A, FIG. 3B and FIG. 3C).

5.1.2.3 SMM and Sulfur Transport in the Phloem

The data demonstrated that SMM makes a major contribution to reduced sulfur transport in the phloem. SMM can be reconverted to Met *via* the action of HMT, and HMT occurs in seeds (Giovanelli *et al.*, 1980). SMM arriving in the phloem could therefore provide much of the Met needed to synthesize grain proteins. There is good indirect evidence that SMM does this: it is essentially absent from mature grains and so must be metabolized (Grunau *et al.*, 1991;

Pimenta *et al.*, 1998), and conversion to Met is its only likely fate. Accordingly, a scheme is presented in FIG. 4, in which SMM is synthesized in the leaves, moves to the grain, and is there recycled to Met for use in protein synthesis. This scheme does not posit a complete separation in space and time of the halves of the SMM cycle, because leaves express HMT (Mudd *et al.*, 1990; Larina *et al.*, 1994) and seed tissues express MMT (Pimenta *et al.*, 1998). Rather, it suggests that while both organs have the potential to carry out the full cycle, flux is mainly from Met to SMM in leaves and from SMM to Met in seeds. This implies that the SMM cycle is tightly regulated.

5.2 EXAMPLE 2 – CLONING AND CHARACTERIZATION OF A cDNA ENCODING MMT FROM TOBACCO

Following the results obtained in Example 1, the inventors screened additional plant sequences for the presence of genes encoding MMT activity. Using the results obtained from the *W. biflora* studies, the partial MMT cDNA sequence from tobacco was isolated using RT-PCR™ with the following primers:

Forward primer: 5'-CAGGGTTTYGTNGARGAYCA-3' (SEQ ID NO:52)

Reverse primer: 5'-CCAGTAACCACNACYTGNGG-3' (SEQ ID NO:53)

The nucleotide sequence of tobacco MMT partial cDNA is shown below:

Nicotiana tabacum MMT partial cDNA sequence (SEQ ID NO:7)

```
cagggttt
tgtcgaggat caGTTTGGCT TGGGGCTTAT TGCAAGGGCA GTTGAAGAAG
GTATTTCTGT CATAAAGCCA TTGGGCATTA TGATCTTCAA CATGGGAGGC
CGTCCTGGGC AAGGTGTTTG CAAACGGTTA TTTGAGCGCC GTGGTCTTCG
TGTTAACAAG CTCTGGCAAA CTAAAATTCT TCAGGCAGCT GACACTGATA
TATCAGCTCT AGTTGAAATT GAAAAGAGTA GCATGCACCG GTTTGAATTT
TTCATGGGAC TTGTTGGAGA TCAGCCAATA TGTGCTCGAA CAGCATGGGC
TTATGGCAAG GCTGGTGGTC GTATCTCTCA TGCTTTATCT GTGTACAGCT
GTCAACTTCG TCAGCCAAAT CAGGTCAGAA AGATATTTGA GTTCATAAAA
AATGGATTCC ATGATATCAG TAATTCTCTG GATTTGTCAT TTGAGGATGA
TGCAGTAGCA GATGAAAAGA TCCCTTTCCT AGCTTATCTT GCTAGTGTGC
TTAAGGAGAA CTCTGTTTTC CCATATGAAT CACCAGCAGG AAGTAGATGG
TTCCGCAATC TGATTGCTGG CTTTATGAAA ACATACCACC ATTTTCCTCT
TA
```

(The portion of the MMT sequence originating from the forward primer is shown in lower case.)

The *Nicotiana tabacum* MMT partial amino acid sequence (SEQ IDNO:8) is shown below:

QGFVEDQFGLGLIARAVEEGISVIKPLGIMIFNMGGRPGQGVCKRLFERRGLRVNKLW
QTKILQAADTDISALVEIEKSSMHRFEFFMGLVGDQPICARTAWAYGKAGGRISHALS
VYSCQLRQPNQVRKIFEFIKNGFHDISNSLDLSFEDDAVADEKIPFLAYLASVLKENS
VFPYESPAGSRWFRNLIAGFMKTYHHFPL

5.3 EXAMPLE 3 – COMPARISON OF CDNAS ENCODING MMT FROM MAIZE AND *WOLLASTONIA*

A comparison of the open reading frames of the maize cDNA sequence (SEQ ID NO:5) and the *Wollastonia* cDNA sequence (SEQ ID NO:1) encoding MMT is shown in FIG. 5A, FIG. 5B, FIG. 5C, FIG. 5D, FIG. 5E and FIG. 5F. Identified in the figure are regions of the sequence that show a high degree of nucleotide conservation (sequences of greater than or equal to at least 12 contiguous nucleotides Regions of the two sequences having less homology than a 12-contiguous nucleic acid sequence have not been indicated on the figure for reasons of clarity. The two sequences, overall, however showed a 76% identity at the polynucleotide sequence level when using the advanced BLASTN v2.0.9 algorithm, available from the National Center for Biotechnology Information, a division of the National Library of Medicine at the National Institutes of Health (http://www.ncbi.nlm.nih.gov/blast/blast.cgi?Jform=1) (Altschul *et al.*, 1990; Gish and States, 1993; Madden *et al.*, 1996; Altschul *et al.*, 1997; Zhang and Madden, 1997). Identified on the alignment are particular nucleic acid sequences (SEQ ID NO:9 to SEQ ID NO:16) that are identical in the two MMT-encoding plant cDNA sequences.

5.4 EXAMPLE 4 – COMPARISON OF MAIZE AND *WOLLASTONIA* MMT POLYPEPTIDES

A comparison of the amino acid sequences of the maize and *Wollastonia* MMT polypeptides is shown in FIG. 6A, FIG. 6B, FIG. 6C and FIG. 6D. Identified in the figure are regions of the sequence that show a high degree of amino acid conservation (sequences of greater than or equal to at least 8 contiguous amino acids). Regions of the two sequences having less homology than an 8-contiguous amino acid sequence have not been indicated on the figure for reasons of clarity. The two sequences, overall, however showed a 62% identity at the amino acid

level when using the advanced BLASTP v2.0.9 algorithm (see Example 3). Identified on the alignment are particular amino acid sequences (SEQ ID NO:26 to SEQ ID NO:44) that are identical in the two plant MMT polypeptides.

5.5 Example 5 – Comparison of cDNAs Encoding MMT from Tobacco and Corn

A comparison of the tobacco partial DNA sequence and the maize cDNA sequence is shown in FIG. 7A and FIG. 7B. Identified in the figure are regions of the sequence that show a high degree of nucleotide conservation (sequences of greater than or equal to at least 14 contiguous nucleotides). Regions of the two sequences having less homology than a 14-contiguous sequence have not been indicated on the figure for clarity. The two sequences, overall, however showed a 78% identity at the polynucleotide sequence level when using the advanced BLASTN v2.0.9 algorithm (see Example 3). Identified on the alignment are particular nucleic acid sequences (SEQ ID NO:21 to SEQ ID NO:25) that are identical in the two cloned genes.

5.6 Example 6 – Comparison of MMT Polypeptides from Tobacco and Corn

A comparison of the tobacco partial amino acid sequence and the maize amino acid sequence of MMT is shown in FIG. 8A and FIG. 8B. Identified in the figure are regions of the sequence that show a high degree of amino acid conservation (sequences of greater than or equal to at least 11 contiguous amino acids). Regions of the two sequences having less homology than an 11-contiguous sequence have not been indicated for reasons of clarity. The two sequences, overall, however showed an 80% identity at the amino acid level when using the advanced BLASTP v2.0.9 algorithm (see Example 3). Identified on the alignment are particular amino acid sequences (SEQ ID NO:45 to SEQ ID NO:51) that are identical in the two plant MMT polypeptides.

6.0 References

The following literature citations as well as those cited above are incorporated in pertinent part by reference herein for the reasons cited in the above text:

U. S. Patent No. 4,332,898, issued Jun. 1, 1982.

U. S. Patent No. 4,342,832, issued Aug. 3, 1982.

U. S. Patent No. 4,356,270, issued Oct. 26, 1982.
U. S. Patent No. 4,362,817, issued Dec. 7, 1982.
U. S. Patent No. 4,371,625, issued Feb. 1, 1983.
U. S. Patent No. 4,554,101, issued Nov. 19, 1985.
U. S. Patent No. 4,554,101, issued Nov. 19, 1985.
U. S. Patent No. 4,683,195, issued Jul. 28, 1987.
U. S. Patent No. 4,683,202, issued Jul. 28, 1987.
U. S. Patent No. 4,873,192, issued Oct. 10, 1989.
U. S. Patent No. 4,943,674, issued Jul. 24, 1990.
U. S. Patent No. 4,945,050, issued Jul. 31, 1990.
U. S. Patent No. 4,965,188, issued Oct. 23, 1990.
U. S. Patent No. 5,097,025, issued Mar. 17, 1992.
U. S. Patent No. 5,106,739, issued Apr. 21, 1992.
U. S. Patent No. 5,110,732, issued May 5, 1992.
U. S. Patent No. 5,139,954, issued Aug. 19, 1992.
U. S. Patent No. 5,176,995, issued Oct. 15, 1991.
U. S. Patent No. 5,177,011, issued Jan. 5, 1993.
U. S. Patent No. 5,240,855, issued Aug. 31, 1993.
U. S. Patent No. 5,268,463, issued Dec. 7, 1993.
U. S. Patent No. 5,322,783, issued Jun. 21, 1994.
U. S. Patent No. 5,324,253, issued Jun. 28, 1994.
U. S. Patent No. 5,324,646, issued Jun. 28, 1994.
U. S. Patent No. 5,378,619, issued Jan. 3, 1995.
U. S. Patent No. 5,399,680, issued Mar. 21, 1995.
U. S. Patent No. 5,401,836, issued Mar. 28, 1995.
U. S. Patent No. 5,405,765, issued Apr. 11, 1995.
U. S. Patent No. 5,441,884, issued Aug. 15, 1995.
U. S. Patent No. 5,442,052, issued Aug. 15, 1995.
U. S. Patent No. 5,447,858, issued Sep. 5, 1995.
U. S. Patent No. 5,459,252, issued Oct. 17, 1995.
U. S. Patent No. 5,466,785, issued Nov. 14, 1995.
U. S. Patent No. 5,472,869, issued Dec. 5, 1995.

U. S. Patent No. 5,484,956, issued Jan. 16, 1996.

U. S. Patent No. 5,491,288, issued Feb. 13, 1996.

U. S. Patent No. 5,504,200, issued Apr. 2, 1996.

U. S. Patent No. 5,530,196, issued Jun. 25, 1996.

U. S. Patent No. 5,538,879, issued Jul. 23, 1996.

U. S. Patent No. 5,539,082, issued Jul. 23, 1996.

U. S. Patent No. 5,563,055, issued Oct. 8, 1996.

U. S. Patent No. 5,569,597, issued Oct. 29, 1996.

U. S. Patent No. 5,576,198, issued Nov. 19, 1996.

U. S. Patent No. 5,589,583, issued Dec. 31, 1996.

U. S. Patent No. 5,589,610, issued Dec. 31, 1996.

U. S. Patent No. 5,589,614, issued Dec. 31, 1996.

U. S. Patent No. 5,595,896, issued Jan. 21, 1997.

U. S. Patent No. 5,604,121, issued Feb. 18, 1997.

U. S. Patent No. 5,608,142, issued Mar. 4, 1997.

U. S. Patent No. 5,608,144, issued Mar.4, 1997.

U. S. Patent No. 5,608,149, issued Mar. 4, 1997.

U. S. Patent No. 5,610,042, issued Mar. 11, 1997.

U. S. Patent No. 5,614,399, issued Mar. 25, 1997.

U. S. Patent No. 5,629,183, issued May 13, 1997.

U. S. Patent No. 5,633,363, issued May 27, 1997.

U. S. Patent No. 5,633,439, issued May 27, 1997.

U. S. Patent No. 5,633,440, issued May 27, 1997.

U. S. Patent No. 5,633,441, issued May 27, 1997.

U. S. Patent No. 5,646,333, issued Jul. 8, 1997.

U. S. Patent No. 5,659,124, issued Aug. 19, 1997.

U. S. Patent No. 5,689,040, issued Nov. 18, 1997.

U. S. Patent No. 5,689,049, issued Nov. 18, 1997.

U. S. Patent No. 5,689,051, issued Nov. 18, 1997.

U. S. Patent No. 5,689,056, issued Nov. 18, 1997.

U. S. Patent No. 5,712,112, issued Jan. 27, 1998.

U. S. Patent No. 5,719,262, issued Feb. 17, 1998.

U. S. Patent No. 5,736,336, issued Apr. 17, 1998.

U. S. Patent No. 5,766,855, issued Jun. 16, 1998.

U. S. Patent No. 5,773,571, issued Jun. 30, 1998.

U. S. Patent No. 5,786,461, issued Jul. 28, 1998.

U. S. Patent No. 5,859,347, issued Jan. 12, 1999.

U. S. Patent No. 5,866,793, issued Feb. 2, 1999.

U. S. Patent No. 5,874,265, issued Feb. 23, 1999.

U. S. Patent No. 5,874,626, issued Feb. 23, 1999.

U. S. Patent No. 5,886,244, issued Mar. 23, 1999.

European Patent Application Publication No. 75,444.

European Patent Application Publication No. 120 516.

European Patent Application Publication No. A0359472.

Japanese Patent No. 1993207886-A4.

Japanese Patent No. 1994062866-A1.

Japanese Patent No. 1997070291-A.

Japanese Patent No. 1987232392-A1.

Japanese Patent No. 1986076298.

PCT Publication No. WO 84/02913.

PCT Publication No. W0 91/16432.

PCT Publication No. WO 96/19256.

Abdullah *et al.*, *Biotechnology*, 4:1087, 1986.

Altschul *et al.*, "Basic local alignment search tool," *J. Mol. Biol.*, 215:403-410, 1990.

Altschul *et al.*, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res.*, 25:3389-3402, 1997.

An *et al.*, *EMBO J.*, 4:277-287, 1985.

Arz *et al.*, *Biochim. Biophys. Acta,* 1218(3):447-452, 1994.

Bagdasarian, Lurz, Ruckert, Franklin, Bagdasarian, Frey, Timmis, "Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in Pseudomonas," *Gene*, 16(1-3):237-247, 1981.

Ballas *et al.*, *Nucleic Acids Res.*, 17:7991-7903, 1989.

Baum *et al.*, *Appl. Environ. Microbiol.*, 56:3420-3428, 1990.

Baumlein, Boerjan, Nagy, Panitz, Inze, Wobus, "Upstream sequences regulating legumin gene expression in heterologous transgenic plants," *Mol. Gen. Genet.*, 225(1):121-128, 1991.

Benbrook *et al.*, *In: Proceedings Bio Expo 1986*, Butterworth, Stoneham, MA, pp. 27-54, 1986.

Berna and Bernier, "Regulated expression of a wheat germin gene in tobacco: oxalate oxidase activity and apoplastic localization of the heterologous protein," *Plant Mol. Biol.*, 33(3):417-429, 1997.

Bevan *et al.*, *Nature*, 304:184, 1983.

Bezzubov *et al.*, *Biokhim. Mikrobiol.*, 28:423-429, 1992.

Bick *et al.*, *Proc. Natl. Acad. Sci. USA*, 95(14):8404-8409, 1998.

Boffa, Carpaneto, Allfrey, *Proc. Natl. Acad. Sci. USA*, 92:1901-1905, 1995.

Bork *et al.*, *Gene*, 212(1):147-153, 1998.

Bork *et al.*, *Plant Physiol.*, 115:864, 1997.

Boronat, Martinez, Reina, Puigdomenech, Palau, "Isolation and sequencing of a 28 kd gluteline-2 gene from maize: Common elements in the 5' flanking regions among zein and glutelin genes," *Plant Sci.*, 47:95-102, 1986.

Brakhage iet al., Biochimie, *72(10):725-734, 1990.*

Brander *et al. Plant Physiol.*, 108:1748, 1995.

Bruehl *et al.*, *Biochim. Biophys. Acta*, 1295:119-124, 1996.

Bussey *et al.*, *Nature*, 387(6632 Suppl.):103-105, 1997.

Bytebier *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:5345-5349, 1987.

Callis, Fromm, Walbot, "Introns increase gene expression in cultured maize cells," *Genes Devel.*, 1:1183-1200, 1987.

Campbell, *In: Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology*, Burden and Von Knippenberg, Eds., Vol. 13, pp. 75-83, Elsevier, Amsterdam, 1984.

Capecchi, "High efficiency transformation by direct microinjection of DNA into cultured mammalian cells," *Cell*, 22(2):479-488, 1980.

Cheng *et al.*, *Proc. Natl. Acad. Sci. USA*, 95(6):2767-2772, 1998.

Christensen *et al.*, *Plant Mol. Biol.*, 12:619-632, 1989.

Christensen *et al.*, *Plant Mol. Biol.*, 18:675-689, 1992.

Christensen, Sharrock, Quail, "Maize polyubiquitin genes: Structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation," *Plant Mol. Biol.*, 18:675-689, 1992.

Christou and Ford, *Annals of Botany*, 75:407-413, 1995.

Christou *et al.*, *Plant Physiol.*, 87:671-674, 1988.

Clepet *et al.*, *Mol. Microbiol.*, 6(21):3109-3119, 1992.

CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California, USA.

Conway and Wickens, *In: RNA Processing*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 40, 1988.

Corey, *Trends Biotechnol.*, 15(6):224-229, 1997.

Corpet, *Nucleic Acids Res.*, 16:10881-10890, 1988.

Cristou *et al.*, *Plant Physiol*, 87:671-674, 1988.

Crossway *et al.*, *Biotechniques*, 4:320-334, 1986.

Curiel, Agarwal, Wagner, Cotten, "Adenovirus enhancement of transferrin-polylysine-mediated gene delivery," *Proc. Natl. Acad. Sci. USA*, 88(19):8850-8854, 1991.

Daniel *et al.*, *J. Mol. Biol.*, 232(2):468-483, 1993.

Danos and Heard, "Recombinant retroviruses as tools for gene transfer to somatic cells," *Bone Marrow Transplant*, 9(Suppl.1):131-138, 1992.

Datta *et al.*, *Biotechnology*, 8:736-740, 1990.

Dayhoff *et al.*, In: *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington, D.C., 1978.

De Wet *et al.*, In: *The Experimental Manipulation of Ovule Tissues*, Chapman *et al.*, Eds., Longman, New York, pp. 197-209, 1985.

Dean *et al.*, *Nucl. Acids Res.*, 14(5):2229-2240, 1986.

Dennis, Gerlach, Pryor, Bennetzen, Inglis, Llewellyn, Sachs, Ferl, Peackocock, "Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize," *Nucl. Acids Res.*, 12:3983-4000, 1984.

D'Halluin *et al.*, *Plant Cell*, 4:1495-1505, 1992.

Eglitis and Anderson, "Retroviral vectors for introduction of genes into mammalian cells," *Biotechniques*, 6(7):608-614, 1988.

Eichel *et al.*, *Eur. J. Biochem.*, 230(3):1053-1058, 1995.

Eichholtz *et al.*, "Expression of mouse dihydrofolate reductase gene confers methotrexate resistance in transgenic petunia plants," *Somat. Cell Mol. Genet.*, 13(1):67-76, 1987.

Faktor, Kooter, Dixon, Lamb, "Functional dissection of a bean chalcone synthase gene promoter in transgenic tobacco plants reveals sequence motifs essential for floral expression," *Plant Mol. Biol.*, 32(5):849-859, 1996.

Ficker, Kirch, Eijlander, Jacobsen, Thompson, "Multiple elements of the S2-RNase promoter from potato (*Solanum tuberosum* L.) are required for cell type-specific expression in transgenic potato and tobacco," *Mol. Gen. Genet.*, 257(2):132-142, 1998.

Finer and McMullen, *In Vitro Cell Dev. Biol.*, 27P: 175-182, 1991.

Fraley *et al.*, *Bio/Technology*, 3:629-635, 1985.

Fraley, Rogers, Horsch, Sanders, Flick, Adams, Bittner, Brand, Fink, Fry, Galluppi, Goldberg, Hoffmann, Woo, "Expression of bacterial genes in plant cells," *Proc. Natl. Acad. Sci. USA*, 80(15):4803-4087, 1983.

French, Janda, Ahlquist, "Bacterial gene inserted in an engineered RNA virus: efficient expression in monocotyledonous plant cells," *Science*, 231:1294-1297, 1986.

Fromm *et al.*, *Biotechnology*, 8:833-839, 1990.

Fromm, Taylor, Walbot, "Expression of genes transferred into monocot and dicot plant cells by electroporation," *Proc. Natl. Acad. Sci. USA*, 82(17):5824-5828, 1985.

Fujimura *et al.*, *Plant Tissue Cult. Lett.*, 2:74, 1985.

Fujioka, *Int. J. Biochem.*, 24:1917-1924, 1992.

Gallie and Young, "The regulation of expression in transformed maize aleurone and endosperm protoplasts," *Plant Physiol.*, 106:929-939, 1994.

Gallie, Feder, Schimke, Walbot, "Post-transcriptional regulation in higher eukaryotes: the role of the reporter gene in controlling expression," *Mol. Gen. Genet.*, 228:258-264, 1991.

Gallie, Lucas, Walbot, "Visualizing mRNA expression in plant protoplasts: factors influencing efficient mRNA uptake and translation," *Plant Cell*, 1:301-311, 1989.

Gallie, Sleat, Turner, Wilson, "Mutational analysis of the tobacco mosaic virus 5'-leader for altered ability to enhance translation," *Nucl. Acids Res.*, 16:883-893, 1988.

Gallie, Sleat, Watts, Turner, Wilson, "A comparison of eukaryotic viral 5'-leader sequences as enhancers of mRNA expression *in vivo*," *Nucl. Acids Res.*, 15:8693-8711, 1987b.

Gallie, Sleat, Watts, Turner, Wilson, "The 5′-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts *in vitro* and *in vivo*," *Nucl. Acids Res.*, 15:3257-3273, 1987a.

GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA.

Gefter, Margulies, Scharff, "A simple method for polyethylene glycol-promoted hybridization of mouse myeloma cells," *Somat. Cell Genet.*, 3(2):231-236, 1977.

Gehrke, Auron, Quigley, Rich, Sonenberg, "5′-Conformation of capped alfalfa mosaic virus ribonucleic acid 4 may reflect its independence of the cap structure or of cap-binding protein for efficient translation," *Biochemistry*, 22:5157-5164, 1983.

Gil and Proudfoot, "A sequence downstream of AAUAAA is required for rabbit beta-globin mRNA 3′-end formation," *Nature*, 312(5993):473-474, 1984.

Giovanelli *et al.*, In: *The Biochemistry of Plants*, Academic Press, New York, 5:453-505, 1980.

Gish and States, "Identification of protein coding regions by database similarity search," *Nature Genet.*, 3:266-272, 1993.

Goding, "Monoclonal Antibodies: Principles and Practice," pp. 60-74. 2nd Edition, Academic Press, Orlando, FL, 1986.

Goelet, Lomonossoff, Butler, Akam, Gait, Karn, "Nucleotide sequence of tobacco mosaic virus RNA," *Proc. Natl. Acad. Sci. USA*, 79:5818-5822, 1982.

Good and Nielsen, *Antisense Nucleic Acid Drug Dev.*, 7(4):431-437, 1997.

Gothel *et al.*, *Eur. J. Biochem.*, 244(1):59-65, 1997.

Graham and van der Eb, "Transformation of rat cells by DNA of human adenovirus 5," *Virology*, 54(2):536-539, 1973.

Graham, Craig, Waterhouse, "Expression patterns of vascular-specific promoters ROIC and Sh in transgenic potatoes and their use in engineering PLRV-resistant plants," *Plant Mol. Biol.*, 33(4):729-735, 1997.

Grosset, Alary, Gautier, Menossi, Martinez-Izquierdo, Joudrier, "Characterization of a barley gene coding for an alpha-amylase inhibitor subunit (Cmd protein) and analysis of its promoter in transgenic tobacco plants and in maize kernels by microprojectile bombardment," *Plant Mol. Biol.*, 34(2):331-338, 1997.

Grunau et al., *J. Plant Nutr.*, 14:653-662, 1991.

Guerineau *et al.*, *Mol. Gen. Genet.*, 262:141-144, 1991.

Gutierrez-Marcos *et al., Proc. Natl. Acad. Sci. USA,* 93:13377-13382, 1996.

Hanvey *et al., Science,* 258:1481-1485, 1992.

Harlow and Lane, *In: Antibodies: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1988.

Hell, *Planta,* 202:138, 1997.

Herrera-Estrella *et al., Nature,* 303:209, 1983.

Hess, *Intern Rev. Cytol.,* 107:367, 1987.

Hesse *et al.,* "Isolation of cDNAs encoding cytosolic and plastidic cysteine synthase isoforms from *Solanum tuberosum,*" *Plant Physiol.,* 116:1604, 1998.

Hibler *et al.,* "Biolistic transformation of Botryotinia fuckeliana," *Curr. Genet.,* 25(2):124-127, 1994.

Higgins *et al., Gene,* 73:237-244, 1988.

Higgins *et al., CABIOS,* 5:151-153, 1989.

Hipp *et al., Microbiol.,* 143(Pt9):2891-2902, 1997.

Hockema, *In: The Binary Plant Vector System,* Offset-durkkerij, Kanters B.V., Alblasserdam, Chapter 5, 1985.

Holsters *et al., Mol. Gen. Genet.,* 163:181-187, 1978.

Hooykaas-Van Slogteren *et al.,* "Expression of Ti plasmid genes in monocotyldeonous plants infected with *Agrobacterium tumefaciens.* 1984.," *Biotechnology,* 24:382-383, 1992b.

Hooykaas-Van Slogteren *et al., Nature,* 311:763-764, 1992a.

Horsch *et al., Science,* 227:1229-1231, 1985.

Huang *et al., Computer Applications in the Biosciences,* 8:155-165, 1992.

Huang, An, McDowell, McKinney, Meagher, "The *Arabidopsis* ACT11 action gene is strongly expressed in tissues of the emerging inflorescence, pollen and developing ovules," *Plant Mol. Biol.,* 33(1):125-139, 1997.

Hudspeth and Grula, "Structure and expression of the maize gene encoding the phosphoenolpyruvate carboxylase isozyme involved in C4 photosynthesis," *Plant Mol. Biol.,* 12:579-589, 1989.

Hummerjohann *et al., Microbiol.,* 144(Pt5):1375-1386, 1998.

Ingelbrecht, Herman, Dekeyser, Van Montagu, Depicker, "Different 3′ end regions strongly influence the level of gene expression in plant cells," *Plant Cell,* 1:671-680, 1989.

Innis *et al.*, Eds. In: *PCR Protocols, A Guide to Methods and Applications*, Academic Press, New York, 1990.

Inoue *et al.*, *Biochim. Biophys. Acta*, 1395(3):315-320, 1998.

Jain *et al.*, *Plant Physiol.*, 105:771-772, 1994.

James *et al.*, *J. Biol. Chem.*, 270:22344-22350, 1995a.

James *et al.*, *Plant Physiol.*, 108:1439-1448, 1995b.

Jameson and Wolf, "The Antigenic Index: A Novel Algorithm for Predicting Antigenic Determinants," *Compu. Appl. Biosci.*, 4(1):181-6, 1988.

Jobling and Gehrke, "Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence," *Nature*, 325:622-625, 1987.

Jones, Dean, Gidoni, Gilbert, Bond-Nutter, Lee, Bedbrook, Dunsmuir, "Expression of bacterial chitinase protein in tobacco leaves using two photosynthetic gene promoters," *Mol. Gen. Genet.*, 212:536-542, 1988.

Jorgensen *et al.*, *Mol. Gen. Genet.*, 207:471, 1987.

Joshi *et al.*, *Nucleic Acids Res.*, 15:9627-9639, 1987.

Joshi *et al.*, *Plant Mol. Biol.*, 37:663-674, 1998.

Joshi, "An inspection of the domain between putative TATA box and translation start site in 79 plant genes," *Nucl. Acids Res.*, 15:6643-6653, 1987.

Kaeppler *et al.*, *Plant Cell Reports*, 9:415-418, 1990.

Kaeppler *et al.*, *Theor. Appl. Genet.*, 84:560-566, 1992.

Kaiser and Kezdy, *Science*, 223:249-255, 1984.

Kim and Leustek, "Cloning and analysis of the gene for cystathionine gamma-synthase grom *Arabidopsis thaliana*," *Plant Mol. Biol.*, 32(6):1117-1124, 1996.

King et al., *Plant Physiol.*, 53:96-103, 1974.

Klee *et al.*, *Bio/Technology*, 3:637-642, 1985.

Klein *et al.*, *Biotechnology*, 6:559-563, 1988a.

Klein *et al.*, *Nature*, 327:70, 1987.

Klein *et al.*, *Plant Physiol.*, 91:440-444, 1988b.

Klein *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:4305-4309, 1988c.

Kohler and Milstein, "Continuous cultures of fused cells secreting antibody of predefined specificity," *Nature*, 256(5517):495-497, 1975.

Kohler and Milstein, "Derivation of specific antibody-producing tissue culture and tumor lines by cell fusion," *Eur. J. Immunol.*, 6(7):511-519, 1976.

Korch *et al.*, *Mol. Gen. Genet.*, 229(1):96-108, 1991.

Kozak, *Nature*, 308:241-246, 1984.

Koziel *et al.*, *Plant Mol. Biol.*, 32:393-405, 1996.

Koziel, Beland, Bowman, Carozzi, Crenshaw, Crossland, Dawson, Desai, Hill, Kadwell, Launis, Lewis, Maddox, McPherson, Meghji, Merlin, Rhodes, Warren, Wright, Evola, "Field performance of elite transgenic maize plants expressing an insecticidal protein derived from *Bacillus thuringiensis*, *Bio/technology*, 11:194-200, 1993.

Krone *et al.*, *FEBS Lett.*, 260(1):6-9, 1990.

Krone *et al.*, *Mol. Gen. Genet.*, 225(2):314-319, 1991.

Kuby, *Immunology 2nd Edition*, W. H. Freeman & Company, NY, 1994.

Kunkel *et al.*, *Methods Enzymol.*, 154:367-382, 1987.

Kunkel, *Proc. Natl. Acad. Sci. USA*, 82:488-492, 1985.

Kurvari *et al.*, *Plant Mol. Biol.*, 29:1235-1252, 1995.

Kyozuka *et al.*, "Anaerobic induction and tissue-specific expression of maize Adh1 promoter in transgenic rice plants and their progeny," *Mol. Gen. Genet.*, 228(1-2):40-48, 1991.

Kyte and Doolittle, *J. Mol. Biol.*, 157:105-132, 1982.

Larina *et al.*, *J. Plant Physiol.*, 41:358-363, 1994.

Last *et al.*, *Theor. Appl. Genet.*, 81:581-588, 1991.

Laue *et al.*, *J. Bacteriol.*, 176:3723-3729, 1994.

Lee *et al.*, *Biochem. Biophys. Res. Commun.*, 247:171-175, 1998.

Leustek *et al.*, *Plant Physiol.*, 105:897-902, 1994.

Li *et al.*, *Plant Cell Reports*, 12:250-255, 1993.

Lorz *et al.*, *Mol. Gen. Genet.*, 199:178, 1985.

Luehrsen and Walbot, "Intron enhancement of gene expression and the splicing efficiency of introns in maize cells," *Mol. Gen. Genet.*, 225:81-93, 1991.

Luehrsen, In: *The Maize Handbook*, Springer-Verlag New York, Inc., New York, NY, 1994.

Luo et al., Plant Mol. Biol. Report., 6:165, 1988.

Lutcke, Chow, Mickel, Moss, Kern, Scheele, "Selection of AUG initiation codons differs in plants and animals," *EMBO J.*, 6:43-48, 1987.

Maas, Laufs, Grant, Korfhage, Werr, "The combination of a novel stimulatory element in the first exon of the maize *shrunken-1* gene with the following intron enhances reporter gene expression 1000-fold," *Plant Mol. Biol.*, 16:199-207, 1991.

Madden *et al.*, "Applications of network BLAST server," *Meth. Enzymol.*, 266:131-141, 1996.

Maddock *et al.*, *Third International Congress of Plant Molecular Biology*, Abstract 372, 1991.

Maloy *et al.*, *In: Microbial Genetics*, 2nd Edition. Jones and Bartlett Publishers, Boston, MA, 1994.

Marcotte *et al.*, *Nature*, 335:454, 1988.

Mascerenhas, Mettler, Pierce, Lowe, "Intron mediated enhancement of heterologous gene expression in maize," *Plant Mol. Biol.*, 15:913-920, 1990.

McCabe *et al.*, *Biotechnology*, 6:923-926, 1988.

McCormick *et al.*, *Plant Cell Reports*, 5:81-84, 1986.

McDevitt, Imperiale, Ali, Nevins, "Requirement of a downstream sequence for generation of a poly(A) addition site," *Cell*, 37(3):993-999, 1984.

McElroy *et al.*, *Plant Cell*, 2:163-171, 1990.

McElroy, Zhang, Cao, Wu, "Isolation of an efficient promoter for use in rice transformation," *Plant Cell*, 2(2):163-171, 1990.

Mogen *et al.*, *Plant Cell*, 2:1261-1272, 1990.

Mollegaard, Buchardt, Egholm, Nielsen, "Peptide nucleic acid. DNA strand displacement loops as artificial transcription promoters," *Proc. Natl. Acad. Sci. USA*, 91(9):3892-3895, 1994.

Mudd *et al.*, *Plant Physiol.*, 93:623-630, 1990.

Munroe *et al.*, *Gene*, 91:151-158, 1990.

Murillo *et al.*, *Arch. Biochem. Biophys.*, 323(1):195-204, 1995.

Murray *et al.*, *Nucleic Acids Res.*, 17:477-498, 1989.

Nakabachi *et al.*, *Insect Biochem. Mol. Biol.*, 27:1057-1062, 1997.

Nawrath, Poirier, Somerville, "Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of *Arabidopsis thaliana* results in high levels of polymer accumulation," *Proc. Natl. Acad. Sci. USA*, 91:12760-12764, 1994.

Needleman *et al.*, *J. Mol. Biol.*, 48:443, 1970.

Neuhaus *et al.*, *Theor. Appl. Genet.*, 75:30, 1987.

Nielsen *et al.*, "Peptide nucleic acids (PNAs): potential antisense and anti-gene agents," *Anticancer Drug Des.*, 8(1):53-63, 1993b.

Nielsen, Egholm, Berg, Buchardt, "Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide," *Science*, 254(5037):1497-1500, 1991.

Oard, Paige, Dvorak, "Chimeric gene expression using maize intron in cultured cells of breadwheat," *Plant Cell. Rep.*, 8:156-160, 1989.

Odell *et al., Nature*, 313:810-812, 1985.

Okumura *et al., Microbiol.*, 143(Pt9):2883-2890, 1997.

Olson, Porter, Rubinstein, Silver, "Mercuric reductase enzyme from a mercury-volatilizing strain of Thiobacillus ferrooxidans," *J. Bacteriol.*, 151(3):1230-1236, 1982.

Omirulleh, Abraham, Golovkin, Stefanov, Karabaev, Mustardy, Morocz, Dudits, "Activity of a chimeric promoter with the doubled CaMV 35S enhancer element in protoplast-derived cells and transgenic plants in maize," *Plant Molecular Biology*, 21(3):415-428, 1993.

Omori *et al., J. Bacteriol.*, 175(3):785-794, 1993.

Orum, Nielsen, Egholm, Berg, Buchardt, Stanley, "Single base pair mutation analysis by PNA directed PCR™ clamping," *Nucl. Acids Res.*, 21(23):5332-5336, 1993.

Orum, Nielsen, Jorgensen, Larsson, Stanley, Koch, "Sequence-specific purification of nucleic acids by PNA-controlled hybrid selection," *BioTechniques*, 19(3):472-480, 1995.

Osjoda *et al., Nature Biotechnology*, 14:745-750, 1996.

Paquet et al., *Can. J. Bot.*, 73:1889-1896, 1995.

Paszkowski *et al., EMBO J.*, 3:2717-2722, 1984.

Pearson *et al., Proc. Natl. Acad. Sci. USA*, 85:2444, 1988.

Pena *et al., Nature*, 325:274, 1987.

Perlak *et al., Proc. Natl. Acad. Sci. USA*, 88:3324-3328, 1991.

Perlak, Deaton, Armstrong, Fuchs, Sims, Greenplate, Fischhoff, "Insect resistant cotton plants," *Bio/technology*, 8:939-943, 1990.

Perlak, Fuchs, Dean, McPherson, Fischhoff, "Modification of the coding sequence enhances plant expression of insect control protein genes," *Proc. Natl. Acad. Sci. USA*, 88:3324-3328, 1991.

Perlak, Stone, Muskopf, Peterson, Parker, McPherson, Wyman, Love, Reed, Biever, Fischhoff, "Genetically improved potatoes: protection from damage by Colorado potato beetles," *Plant Mol. Biol.*, 22:313-321, 1993.

Perry-O'Keefe, Yao, Coull, Fuchs, Egholm, "Peptide nucleic acid pre-gel hybridization: an alternative to southern hybridization," *Proc. Natl. Acad. Sci. USA*, 93(25):14670-14675, 1996.

Person *et al., Methods of Molecular Biology,* 24:307-331, 1994.

Pimenta et al., *Plant Physiol.,* 118:431-438, 1998.

Poogin and Skryabin, "The 5' untranslated leader sequence of potato virus X RNA enhances the expression of the heterologous gene *in vivo*," *Mol. Gen. Genet.*, 234:329-331, 1992.

Potrykus, Paszkowski, Saul, Petruska, Shillito, "Molecular and general genetics of a hybrid foreign gene introduced into tobacco by direct gene transfer," *Mol. Gen. Genet.*, 199(2):169-177, 1985.

Pott *et al., Microbiol.,* 144(Pt7):1881-1894, 1998.

Powles *et al., Microbiol.,* 142(Pt9):2543-2548, 1996.

Prokop and Bajpai, "Recombinant DNA Technology I," *Ann. N.Y. Acad. Sci.,* Volume 646, 1991.

Proudfoot, *Cell,* 64:671-674, 1991.

Rathinasabapathi et al., *Proc. Natl. Acad. Sci. USA,* 94:3454-3458, 1997.

Ravanel *et al., Plant Mol. Biol.,* 29(4):875-882, 1995.

Ravanel *et al., Proc. Natl. Acad. Sci. USA,* 95(13):7805-7812, 1998.

Riggs *et al., Proc. Natl. Acad. Sci. USA,* 83:5602-5606, 1986.

Rogers *et al., In: Methods For Plant Molecular Biology*, A. Weissbach and H. Weissbach, eds., Academic Press Inc., San Diego, CA 1988.

Russell and Fromm, "Tissue-specific expression in transgenic maize for four endosperm promoters from maize and rice," *Transgenic Res.*, 6(2):157-168, 1997.

Sadofsky and Alwine, "Sequences on the 3' side of hexanucleotide AAUAAA affect efficiency of cleavage at the polyadenylation site," *Molec. Cell. Biol.*, 4(8):1460-1468, 1984.

Saito *et al., J. Biol. Chem.,* 270(27):16321-16326, 1995.

Sambrook *et al.,* In: *Molecular Cloning: A Laboratory Manual,* 2nd Ed.; Cold Spring Harbor Laboratory Press, Plainview, New York, 1989.

Sanfacon *et al., Genes Dev.,* 5:141-149, 1991.

Sanford *et al., Particulate Science and Technology,* 5:27-37, 1987.

Schiffmann *et al.,* "Isolation of cDNA clones encoding adenosine-5'-phosphosulfate-kinase (EC2.7.1.25) from *Catharanthus roseus* and an isoform (akn2) from *Arabidopsis*," *Plant Physiol.,* 117(3):1125, 1998.

Segal, "Biochemical Calculations" 2nd Edition, John Wiley & Sons, New York, 1976.

Setya *et al., Proc. Natl. Acad. Sci. USA,* 93(23):13383-13388, 1996.

Shaw and Kamen, "A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation," *Cell*, 46(5):659-667, 1986.

Shaw and Kamen, *In: RNA Processing*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 220, 1987.

Singh *et al., Theor. Appl. Genet.,* 96:319-324, 1998.

Sleat, Gallie, Jefferson Bevan, Turner, Wilson, "Characterization of the 5'-leader sequence of tobacco mosaic virus RNA as a general enhancer of translation *in vitro*," *Gene*, 217:217-225, 1987.

Sleat, Hull, Turner, Wilson, "Studies on the mechanism of translational enhancement by the 5'-leader sequence of tobacco mosaic virus RNA," *Eur. J. Biochem.*, 175:75-86, 1988.

Smith and Hood, "*Agrobacterium tumefaciens* transformation of monocotyledons," *Crop Science*, 35:301-309, 1995.

Smith *et al., Adv. Appl. Math.*, 2:482, 1981.

Speich *et al., Microbiol.,* 140(Pt6):1273-1284, 1994.

Spielmann *et al., Mol. Gen. Genet.*, 205:34, 1986.

Tanaka, Mita, Ohta, Kyozuka, Shimamoto, Nakamura, "Enhancement of foreign gene expression by a dicot intron in rice but not in tobacco is correlated with an increased level of mRNA and an efficient splicing of the intron," *Nucl. Acids Res.*, 18:6767-6770, 1990.

Thisted, Just, Petersen, Hyldig-Nielsen, Godtfredsen, *Cell Vision*, 3:358-363, 1996.

Thompson et al., *Nucleic Acids. Res.*, 22:4673-4680, 1994.

Tomes *et al.*, "Transgenic Tobacco Plants and their Progeny Derived by Microprojectile Bombardment of Tobacco Leaves," *Plant Mol. Biol.*, 14, 261-268, 1990.

Tomes *et al.*, "Direct DNA transfer into intact plant cells *via* microprojectile bombardment," In: *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, Gamborg and Phillips, Eds., Springer-Veriag, Berlin, 1995.

Topczewski *et al., Curr. Genet.,* 31(4) :348-356, 1997.

Toriyama *et al., Theor Appl. Genet.*, 73:16, 1986.

Treacy, Hattori, Prud'homme, Barbour, Boutilier, Baszczynski, Huang, Johnson, Miki, "Bnm1, a *Brassica* pollen-specific gene," *Plant Mol. Biol.*, 34(4):603-611, 1997.

Trossat *et al., Plant Physiol.,* 111:965-973, 1996.

Uchimiya *et al.*, *Mol. Gen. Genet.*, 204:204, 1986.

Valle *et al.*, *Plant Cell Physiol.*, 39:458- 461, 1998.

Van Camp, Herouart, Willekens, Takahashi, Saito, Van Montagu, Inze, "Tissue-specific activity of two manganese superoxide dismutase promoters in transgenic tobacco," *Plant Physiol.*, 112(2):525-535, 1996.

Vander, Van Montagu, Inze, Boerjan, "Tissue-specific expression conferred by the S-adenosyl-L-methionine synthetase promoter of Arabidopsis thaliana in transgenic poplar," *Plant Cell Physiol.*, 37(8):1108-1115, 1996.

Vasil *et al.*, "Herbicide-resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus," *Biotechnology*, 10:667-674, 1992.

Vasil, *Biotechnology*, 6:397, 1988.

Vasil, Clancy, Ferl, Vasil, Hannah, "Increased gene expression by the first intron of maize *shrunken-1* locus in grass species," *Plant Physiol.*, 91:1575-1579, 1989.

Velten and Schell, "Selection-expression plasmid vectors for use in genetic transformation of higher plants," *Nucl. Acids Res.*, 13(19):6981-6998, 1985.

Velten *et al.*, *EMBO J.*, 3:2723-2730, 1984.

Veselkov, Demidov, Nielsen, Frank-Kamenetskii, "A new class of genome rare cutters," *Nucl. Acids Res.*, 24(13):2483-2487, 1996.

Vickers, Griffith, Ramasamy, Risen, Freier, "Inhibition of NF-kappa B specific transcriptional activation by PNA strand invasion," *Nucl. Acids Res.*, 23(15):3003-3008, 1995.

Wagner, Zatloukal, Cotten, Kirlappos, Mechtler, Curiel, Birnstiel, "Coupling of adenovirus to transferrin-polylysine/DNA complexes greatly enhances receptor-mediated gene delivery and expression of transfected genes," *Proc. Natl. Acad. Sci. USA*, 89(13):6099-6103, 1992.

Walker and Gaastra, Eds., *Techniques in Molecular Biology*, MacMillan Publishing Company, New York, 1983.

Wang *et al.*, "Transient Expression of Foreign Genes in Rice, Wheat and Soybean Cells Following Particle Bombardment," *Plant Mol. Biol.*, 11, 433-439, 1988.

Wang, *J. Am. Chem. Soc.*, 118:7667-7670, 1996.

Weibull *et al.*, *Plant Physiol.*, 92:222-226, 1990.

Weissinger *et al.*, *Ann. Rev. Genet.*, 22:421-477, 1988.

White, *J. Mar. Res.*, 40:529-536, 1982.

Wickens and Stephenson, "Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation," *Science*, 226(4678):1045-1051, 1984.

Wickens *et al.*, *In: RNA Processing*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 9, 1987.

Wilson, Flint, Deaton, Fischhoff, Perlak, Armstrong, Fuchs, Berberich, Parks, Stapp, "Resistance of cotton lines containing a *Bacillus thuringiensis* toxin to pink bollworm (Lepidopteran: Gelechiidae) and other insects," *J. Econ. Entomol.*, 4:1516-1521, 1992.

Wolf *et al.*, "An Integrated Family of Amino Acid Sequence Analysis Programs," *Compu. Appl. Biosci.*, 4(1):187-91, 1988.

Wong and Neumann, "Electric field mediated gene transfer," *Biochim. Biophys. Res. Commun.*, 107(2):584-587, 1982.

Yamada *et al.*, *Plant Cell Rep.*, 4:85, 1986.

Yang *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:4144-4148, 1990.

Yin, Chen, Beachy, "Promoter elements required for phloem-specific gene expression from the RTBV promoter in rice," *Plant J.*, 12(5):1179-1188, 1997a.

Yin, Zhu, Dai, Lamb, Beachy, "RF2a, a bZIP transcriptional activator of the phloem-specific rice tungro bacilliform virus promoter, functions in vascular development," *EMBO J.*, 16(17):5247-5259, 1997b.

Zhang and Madden, "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation," *Genome Res.*, 7:649-656, 1997.

Zhang *et al.*, *Proc. Natl. Acad. Sci. USA*, 94:4504-4509, 1997.

Zhou *et al.*, "Introduction of exogenous DNA into cotton embryos," *Methods Enzymol.*, 101:433-481, 1983.

All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the composition, methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be

substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims. Accordingly, the exclusive rights sought to be patented are as described in the claims below.

CLAIMS:

1. An isolated polynucleotide that:

(a) encodes a polypeptide having S-adenosylmethionine:methionine S-methyltransferase (MMT) activity and that comprises an at least 7 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8;

(b) encodes a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity and at least about 75% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8;

(c) comprises an at least 31 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or

(d) hybridizes to the sequence of from position 82 to 3345 of SEQ ID NO:1 or to the sequence of from position 70 to 3282 of SEQ ID NO:3, or to the sequence of from position 1 to 610 of SEQ ID NO:7, or that hybridizes to the complement thereof, under stringent hybridization conditions.

2. The isolated polynucleotide according to claim 1, comprising a sequence region that encodes a polypeptide having an at least 7 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

3. The isolated polynucleotide according to claim 1 or 2, comprising a sequence region that encodes a polypeptide having an at least 9 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

4. The isolated polynucleotide in accordance with any one of claims 1 to 3, comprising a sequence region that encodes a polypeptide having an at least 11 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

5. The isolated polynucleotide in accordance with any preceding claim, comprising a sequence region that encodes a polypeptide having an at least 13 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

6. The isolated polynucleotide in accordance with any preceding claim, comprising a sequence region that encodes a polypeptide having an at least 15 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

7. The isolated polynucleotide in accordance with any preceding claim, comprising a sequence region that encodes a polypeptide having an at least 17 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

8. The isolated polynucleotide in accordance with any preceding claim, comprising a sequence region that encodes a polypeptide having the sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

9. The isolated polynucleotide in accordance with any preceding claim, comprising a sequence region that encodes a polypeptide having the sequence of SEQ ID NO:2.

10. The isolated polynucleotide in accordance with any one of claims 1 to 8, comprising a sequence region that encodes a polypeptide having the sequence of SEQ ID NO:4.

11. The isolated polynucleotide in accordance with any one of claims 1 to 8, comprising a sequence region that encodes a polypeptide having the sequence of SEQ ID NO:8.

12. The isolated polynucleotide according to claim 1, comprising a sequence region that encodes a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity and at least about 75% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8;

13. The isolated polynucleotide according to claim 12, comprising a sequence region that encodes a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity and at least about 78% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

14. The isolated polynucleotide in accordance with any one of claims 12 to 13, comprising a sequence region that encodes a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity and at least about 80% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8;

15. The isolated polynucleotide in accordance with any one of claims 12 to 14, comprising a sequence region that encodes a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity and at least about 85% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8;

16. The isolated polynucleotide in accordance with any one of claims 12 to 15, comprising a sequence region that encodes a polypeptide having S-adenosylmethionine:methionine S-

methyltransferase activity and at least about 90% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8;

17. The isolated polynucleotide in accordance with any one of claims 12 to 16, comprising a sequence region that encodes a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity and at least about 95% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8;

18. The isolated polynucleotide in accordance with any one of claims 12 to 17, comprising a sequence region that encodes a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity and at least about 98% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8;

19. The isolated polynucleotide according to claim 1, comprising an at least 31 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7.

20. The isolated polynucleotide according to claim 19, comprising an at least 33 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7.

21. The isolated polynucleotide in accordance with any one of claims 19 to 20, comprising an at least 35 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7.

22. The isolated polynucleotide in accordance with any one of claims 19 to 21, comprising an at least 40 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7.

23. The isolated polynucleotide in accordance with any one of claims 19 to 22, comprising an at least 45 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7.

24. The isolated polynucleotide in accordance with any one of claims 19 to 23, comprising an at least 50 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7.

25. The isolated polynucleotide in accordance with any one of claims 19 to 24, comprising an at least 55 contiguous nucleotide sequence from SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7.

26. The isolated polynucleotide in accordance with any one of claims 19 to 25, comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7.

27. The isolated polynucleotide in accordance with any one of claims 19 to 26, comprising the nucleotide sequence of SEQ ID NO:1.

28. The isolated polynucleotide in accordance with any one of claims 19 to 26, comprising the nucleotide sequence of SEQ ID NO:3.

29. The isolated polynucleotide in accordance with any one of claims 19 to 26, comprising the nucleotide sequence of SEQ ID NO:7.

30. The isolated polynucleotide according to claim 1, comprising a sequence region that hybridizes to the sequence of from position 82 to 3345 of SEQ ID NO:1 or to the sequence of from position 70 to 3282 of SEQ ID NO:3, or to the sequence of from position 1 to 610 of SEQ ID NO:7, or to the complement thereof, under stringent hybridization conditions.

31. The isolated polynucleotide according to claim 30, comprising a sequence region that hybridizes to the sequence of from position 82 to 3345 of SEQ ID NO:1 or the complement thereof, under stringent hybridization conditions.

32. The isolated polynucleotide in accordance with any one of claims 30 to 31, comprising a sequence region that hybridizes to the sequence of from position 82 to 3345 of SEQ ID NO:1, or the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

33. The isolated polynucleotide according to claim 30, comprising a sequence region that hybridizes to the sequence of from position 70 to 3282 of SEQ ID NO:3, or the complement thereof, under stringent hybridization conditions.

34. The isolated polynucleotide according to claim 33, comprising a sequence region that hybridizes to the sequence of from position 70 to 3282 of SEQ ID NO:3, or the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

35. The isolated polynucleotide according to claim 30, comprising a sequence region that hybridizes to the sequence of from position 1 to 610 of SEQ ID NO:7, or the complement thereof, under stringent hybridization conditions.

36. The isolated polynucleotide according to claim 35, comprising a sequence region that hybridizes to the sequence of from position 1 to 610 of SEQ ID NO:7, or the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

37. An isolated polynucleotide that comprises:

(a) a sequence region that consists of at least 31 contiguous nucleotides that have the same sequence as, or are complementary to, at least 31 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:7; or

(b) a sequence region of from 31 to about 5000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

38. The isolated polynucleotide according to claim 37, comprising a sequence region that consists of at least 31 contiguous nucleotides that have the same sequence as, or are complementary to, at least 31 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:7; or

39. The isolated polynucleotide in accordance with any one of claims 37 to 38, wherein said polynucleotide is from about 100 to about 10,000 nucleotides in length.

40. The isolated polynucleotide in accordance with any one of claims 37 to 39, wherein said nucleic acid segment is from about 1000 to about 5,000 nucleotides in length.

41. The isolated polynucleotide in accordance with any one of claims 37 to 40, wherein said nucleic acid segment is from about 2000 to about 4,000 nucleotides in length.

42. The isolated polynucleotide according to claim 37, comprising a sequence region of from 31 to about 5000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or that hybridizes to the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

43. The isolated polynucleotide according to claim 42, comprising a sequence region of from 35 to about 5000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or that hybridizes to the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

44. The isolated polynucleotide in accordance with any one of claims 42 to 43, comprising a sequence region of from 40 to about 5000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or that hybridizes to the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

45. The isolated polynucleotide in accordance with any one of claims 42 to 44, comprising a sequence region of from 45 to about 5000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or that hybridizes to the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

46. The isolated polynucleotide in accordance with any one of claims 42 to 45, comprising a sequence region of from 50 to about 5000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or that hybridizes to the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

47. The isolated polynucleotide in accordance with any one of claims 42 to 43, comprising a sequence region of from 31 to about 4000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or that hybridizes to the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

48. The isolated polynucleotide according to claim 47, comprising a sequence region of from 31 to about 3000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or that hybridizes to the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

49. The isolated polynucleotide in accordance with any one of claims 47 to 48, comprising a sequence region of from 31 to about 2000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or that hybridizes to the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

50. The isolated polynucleotide in accordance with any one of claims 47 to 49, comprising a sequence region of from 31 to about 1000 nucleotides in length that hybridizes to the nucleic acid segment of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:7; or that hybridizes to the complement thereof, under hybridization conditions comprising a salt concentration of from about 0.02 M to about 0.15 M, and a temperature of from about 50°C to about 70°C.

51. The polynucleotide of any preceding claim, further defined as an RNA, a PNA, or a DNA segment.

52. The polynucleotide of any preceding claim, comprised within a recombinant vector.

53. The polynucleotide of any preceding claim, comprised within a plasmid, cosmid, phage, phagemid, baculovirus, bacterial artificial chromosome, or yeast artificial chromosome vector.

54. The polynucleotide of any preceding claim, comprised within a recombinant virus or virion.

55. The polynucleotide of any preceding claim, operably linked to a promoter.

56. The polynucleotide of any preceding claim, operably linked to a heterologous promoter.

57. The polynucleotide of any preceding claim, operably linked to a plant-expressible constitutive, inducible, or tissue-specific promoter.

58. The polynucleotide of any preceding claim, operably linked to a plant-expressible promoter selected from the group consisting of corn sucrose synthetase 1, corn alcohol dehydrogenase 1, corn light harvesting complex, corn heat shock protein, pea small subunit RuBP carboxylase, Ti plasmid mannopine synthase, Ti plasmid nopaline synthase, petunia chalcone isomerase, bean glycine rich protein 1, Potato patatin, lectin, CaMV 35S, ALS, ubiquitin, globulin 1, cruciferin, napin, B-conglycinin, phaseolin, gama zein, and the S-E9 small subunit RuBP carboxylase promoter.

59. The polynucleotide of any preceding claim, that encodes a polypeptide that comprises the sequence of any one of SEQ ID NO:26 to SEQ ID NO:51.

60. The polynucleotide of any preceding claim, comprising the sequence of any one of SEQ ID NO:9 to SEQ ID NO:25.

61. A polynucleotide in accordance with any preceding claim for use in preparing a recombinant vector.

62. A polynucleotide in accordance with any one of claims 1 to 60, for use in the preparation of a transgenic plant.

63. A polynucleotide in accordance with any one of claims 1 to 60, for use in screening a plant nucleic acid library to identify a gene encoding a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity.

64. Use of a polynucleotide in accordance with any one of claims 1 to 60 to identify a polynucleotide that encodes a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity.

65. Use of a polynucleotide in accordance with any one of claims 1 to 60 in the preparation of a hybridization probe for screening a plant nucleic acid library to identify a gene encoding a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity.

66. Use of a polynucleotide in accordance with any one of claims 1 to 60 in the generation of a recombinant vector for use in producing a transformed plant cell or plant tissue.

67. Use of a polynucleotide in accordance with any one of claims 1 to 60 in the generation of a pluripotent plant cell that expresses a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity.

68. A host cell comprising a vector having at least a first heterologous expression unit comprising a polynucleotide in accordance with any one of claims 1 to 60.

69. The host cell according to claim 68, wherein said host cell is a bacterial cell.

70. The host cell in accordance with any one of claims 68 to 69, wherein said host cell is an *Escherichia, Salmonella* or *Agrobacterium* cell.

71. The host cell according to claim 68, wherein said cell is an eukaryotic cell.

72. A virus comprising a polynucleotide in accordance with any one of claims 1 to 60.

73. A host cell in accordance with any one of claims 68 to 71 or a virus in accordance with claim 72, for use in the expression of a recombinant polypeptide.

74. Use of a host cell in accordance with any one of claims 68 to 71 or a virus in accordance with claim 72, in the generation of a pluripotent plant tissue that expresses a nucleic acid segment encoding a polypeptide having S-adenosylmethionine:methionine S-methyltransferase activity.

75. An isolated polypeptide encoded by the polynucleotide in accordance with any one of claims 1 to 60.

76. The isolated polypeptide according to claim 75, comprising an at least 7 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

77. The isolated polypeptide in accordance with any one of claims 75 to 76, comprising an at least 9 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

78. The isolated polypeptide in accordance with any one of claims 75 to 77, comprising an at least 11 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

79. The isolated polypeptide in accordance with any one of claims 75 to 78, comprising an at least 13 contiguous amino acid sequence from SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

80. The isolated polypeptide in accordance with any one of claims 75 to 79, comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

81. The isolated polypeptide according to claim 75, that has at least about 75% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

82. The isolated polypeptide according to claim 81, that has at least about 78% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

83. The isolated polypeptide in accordance with any one of claims 81 to 82, that has at least about 80% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

84. The isolated polypeptide in accordance with any one of claims 81 to 83, that has at least about 85% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

85. The isolated polypeptide in accordance with any one of claims 81 to 84, that has at least about 90% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

86. Use of a polypeptide in accordance with any one of claims 81 to 85, in the preparation of an antibody that specifically binds to an S-adenosylmethionine:methionine S-methyltransferase polypeptide.

87. An antibody that specifically binds to a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

88. An immunodetection kit comprising, in suitable container means, an antibody according to claim 87, an immunodetection reagent, and instructions for using said antibody.

89. A nucleic acid detection kit comprising, in suitable container means, at least a first isolated nucleic acid segment comprising the polynucleotide in accordance with any one of claims 1 to 60, a detection reagent, and instructions for using said nucleic acid segment.

90. A recombinant vector comprising the polynucleotide in accordance with any one of claims 1 to 60.

91. A composition comprising the polynucleotide in accordance with any one of claims 1 to 60, the host cell in accordance with any one of claims 68 to 71, the virus according to claim 72, the polypeptide in accordance with any one of claims 75 to 85, or the vector according to claim 90.

92. A transgenic plant comprising:

(a) a heterologous nucleic acid segment that comprises the polynucleotide in accordance with any one of claims 1 to 60;

(b) the host cell in accordance with any one of claims 68 to 71;

(c) the virus according to claim 72;

(d) the polypeptide in accordance with any one of claims 75 to 85; or

(e) the vector according to claim 90.

93. The transgenic plant according to claim 92, having stably incorporated into its genome a heterologous nucleic acid segment that comprises the polynucleotide in accordance with any one of claims 1 to 60, wherein said polynucleotide is operably linked to a promoter that expresses said polynucleotide in said transgenic plant.

94. The transgenic plant in accordance with any one of claims 92 to 93, wherein said plant expresses said heterologous nucleic acid segment to produce a polypeptide that has S-adenosylmethionine:methionine S-methyltransferase activity and at least about 75% sequence identity with the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

95. The transgenic plant in accordance with any one of claims 92 to 94, wherein said plant expresses said heterologous nucleic acid segment to produce a polypeptide that has S-adenosylmethionine:methionine S-methyltransferase activity, wherein said polypeptide has

at least about 78% sequence identity with the sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

96. The transgenic plant in accordance with any one of claims 92 to 95, wherein said plant expresses said heterologous nucleic acid segment to produce a polypeptide that has S-adenosylmethionine:methionine S-methyltransferase activity, wherein said polypeptide has at least about 80% sequence identity with the sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

97. The transgenic plant in accordance with any one of claims 92 to 96, wherein said plant expresses said heterologous nucleic acid segment to produce a polypeptide that has S-adenosylmethionine:methionine S-methyltransferase activity, wherein said polypeptide has at least about 85% sequence identity with the sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

98. The transgenic plant in accordance with any one of claims 92 to 97, wherein said plant expresses said heterologous nucleic acid segment to produce a polypeptide that has S-adenosylmethionine:methionine S-methyltransferase activity, wherein said polypeptide has at least about 90% sequence identity with the sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:8.

99. The transgenic plant in accordance with any one of claims 92 to 98, wherein said plant is a monocotyledonous or a dicotyledonous plant.

100. The transgenic plant in accordance with any one of claims 92 to 99, wherein said plant is a grain, tree, legume, fiber, vegetable, fruit, berry, nut, citrus, grass, cactus, succulent, floral, or ornamental plant.

101. The transgenic plant in accordance with any one of claims 92 to 100, wherein said plant is a corn, rice, millet, tobacco, alfalfa, soybean, bean, sorghum, pea, *Brassica*, safflower, potato, coconut, palm, pumpkin, squash, poppy, sesame, peanut, cocoa, coffee, tomato, flax, canola, sunflower, cotton, flax, kapok, wheat, oat, barley, walnut, pecan, almond, or rye plant.

102. A progeny of any generation of the transgenic plant in accordance with any one of claims 92 to 101.

103. A seed of any generation of the transgenic plant in accordance with any one of claims 92 to 101.

104. A seed of any generation of the progeny according to claim 103.

105. A plant grown from the seed in accordance with any one of claims 103 to 104.

106. A method for detecting an MMT-encoding polynucleotide in a sample, comprising the steps of:

(c) contacting a population of polynucleotides suspected of encoding an S-adenosylmethionine:methionine S-methyltransferase polypeptide with at least a first labeled polynucleotide in accordance with any one of claims 1 to 60, under conditions effective to allow hybridization of substantially complementary nucleic acids; and

(d) detecting the hybridized complementary nucleic acids so formed.

107. A method for detecting an S-adenosylmethionine:methionine S-methyltransferase polypeptide in a biological sample comprising contacting a biological sample suspected of containing said polypeptide with a labeled antibody according to claim 87, under conditions effective to allow the formation of immune complexes, and detecting the immune complexes so formed.

108. A method of increasing the amount of S-adenosylmethionine:methionine S-methyltransferase polypeptide in a plant cell comprising, expressing in said plant cell a biologically effective amount of a polynucleotide in accordance with any one of claims 1 to 60.

109. A method of increasing the level of sulfur amino acids in a plant cell comprising, expressing in said plant cell a biologically effective amount of a polynucleotide in accordance with any one of claims 1 to 60.

110. A method of increasing the level of cysteine or methionine in a plant cell comprising, expressing in said plant cell a biologically effective amount of a polynucleotide in accordance with any one of claims 1 to 60.

111. A method for modulating the biosynthesis of a sulfur compound in a plant, said method comprising the step of (a) transforming said plant with a polynucleotide in accordance with any one of claims 1 to 60 operably linked to a promoter that drives expression in said plant; and (b) growing the plant so transformed under conditions effective to modulate the biosynthesis of said compound in said plant.

112. The method according to claim 111, wherein the biosynthesis of said sulfur compound in said transformed plant is elevated relative to an untransformed plant of the same species.

113. The method in accordance with any one of claims 111 to 112, wherein said sulfur compound is methionine or cysteine.

114. The method in accordance with any one of claims 111 to 113, said method further comprising the steps of (c) growing said transformed plant under conditions effective for obtaining seeds from said plant, and (d) collecting the seeds so produced by said transformed plant.

115. The method in accordance with any one of claims 111 to 114, said method further comprising the step of (e) transforming said plant cell with at least a second polynucleotide that encodes at least one methionine synthesis enzyme polypeptide, wherein said second polynucleotide is operably linked to a promoter capable of expressing said second polynucleotide in said plant cell to produce said methionine synthesis enzyme in said cell.

116. A method for increasing the amount of sulfur-containing amino acids in a plant seed, said method comprising the steps of (a) growing a transgenic plant in accordance with any one of claims 92 to 101, under conditions effective to produce seed in said plant, and (b) obtaining the seed produced from said plant.

117. The method of claim 116, wherein said seed are corn, rice, millet, tobacco, alfalfa, soybean, bean, sorghum, pea, *Brassica*, safflower, potato, tomato, flax, canola, sunflower, cotton, flax, kapok, wheat, oat, barley, or rye seed.

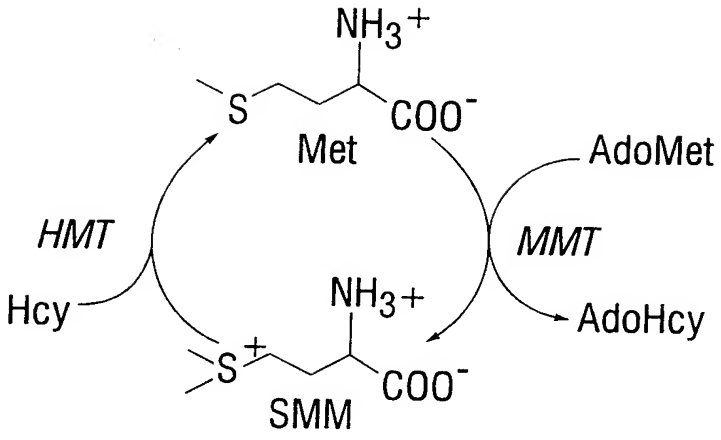
NH3+
S
COO-
Met
AdoMet
HMT
MMT
Hcy
NH3+
AdoHcy
S+
COO-
SMM

FIG. 1

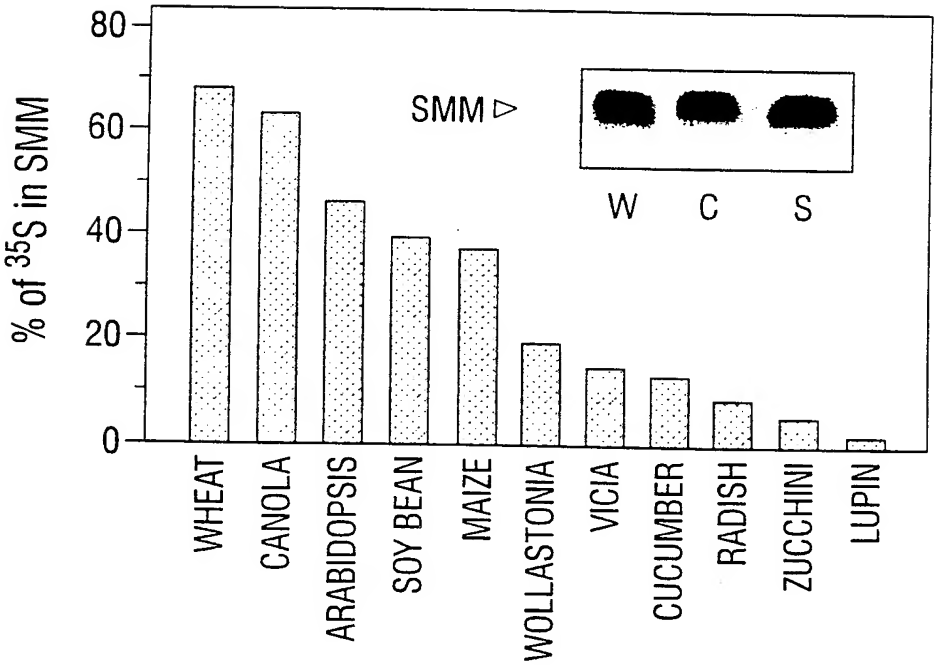
% of 35S in SMM
80
60
40
20
0
SMM
W
C
S
WHEAT
CANOLA
ARABIDOPSIS
SOY BEAN
MAIZE
WOLLASTONIA
VICIA
CUCUMBER
RADISH
ZUCCHINI
LUPIN

FIG. 2A

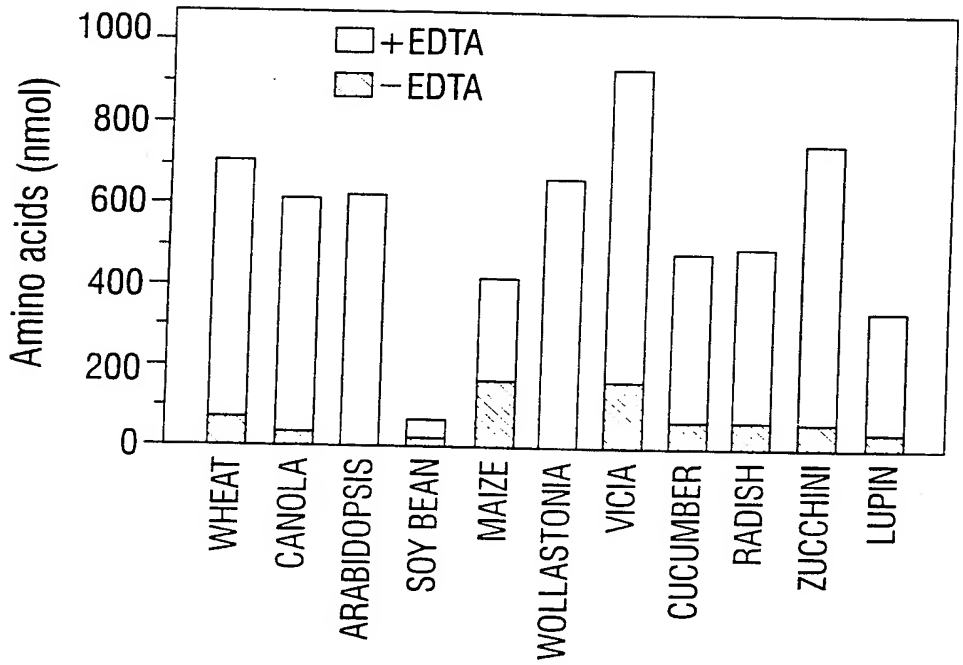
1000
800
600
400
200
0
Amino acids (nmol)
+EDTA
−EDTA
WHEAT
CANOLA
ARABIDOPSIS
SOY BEAN
MAIZE
WOLLASTONIA
VICIA
CUCUMBER
RADISH
ZUCCHINI
LUPIN

FIG. 2B

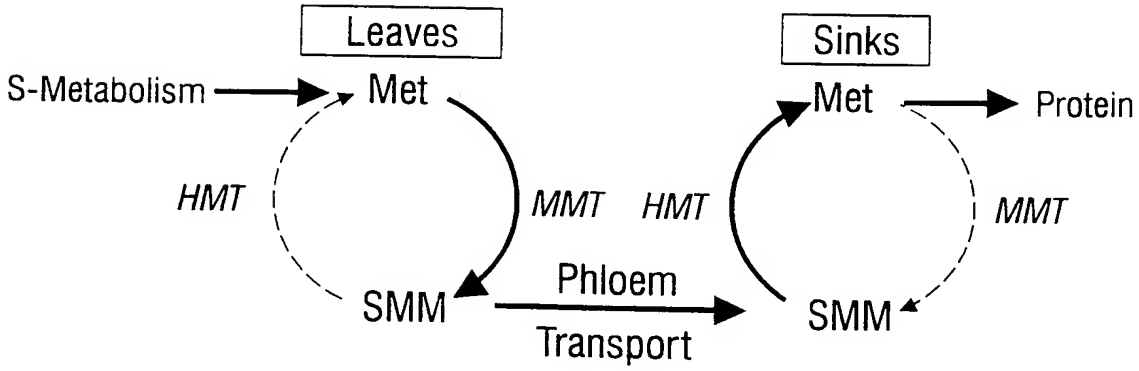
Leaves
Sinks
S-Metabolism
Met
Met
Protein
HMT
MMT
HMT
MMT
Phloem
Transport
SMM
SMM

FIG. 4

```
Zm(    1) MAALAGEDKDVDAFLADCTASGDAAYGAAKAVLERLHAPATRPAARRLLG
At(    1) MADLSS-----VDEFLNQCKQSGDAAYGALRSVLERLEDPNTRSKARIFLS
Wb(    1) MAAVTGLYGSIDEFLNHCSQSGDSAYSALRSLLERLEKPDTRTEARIFLA

Zm(   51) AVRRRFAASRAAGEDCFRTFHFRIHDVVLDP-HVQGFQQMKKLTMMEIPS
At(   47) DIYKRVGSS-----ETSLQTYHFHIQDIYLD--QYEGFQSRKKLTMMVIPS
Wb(   51) HLQKKLDND-GASQRCLETYHFQIQDIYLDRNEGTGYQNRKKFTMMVIPS

Zm(  100) IFIPEDWSFTFYEGLNRHPDSIFRDKTVAELGCGNGWISIALAEKWCPSK
At(   91) IFIPEDWSFTFYEGLNRHPDTIFKDKTVSELGCGNGWISIAIAAKWLPSK
Wb(  100) IFMPEDWSFTFYEGINRHPDSIFKDKTVAELGCGNGWISIAIAEKWLPLK

Zm(  150) VYGLDINPRAVKIAWINLYLNALDDDGLPIYDGEGKTLLDRVEFYESDLL
At(  141) VYGLDINPRAVKISWINLYLNALDDNGEPVYDEEKKTLLDRVEFYESDLL
Wb(  150) VYGLDINPRAVKISWINLYLNAFDEDGQPVYDSESKTLLDRVEFYESDLL

Zm(  200) SYCRDNKIELDRIVGCIPQILNPNPEAMSKIVTENSSEEFLYALSNYCAL
At(  191) GYCRDNKIQLERIVGCIPQILNPNPEAMSKLITENASEEFLHSLSNYCAL
Wb(  200) SYCRDNHIELERIVGCIPQILNPNPDAMSKLVTENASEEFLHSLSNYCAL

Zm(  250) QGFVEDQFGLGLIARAVEEGISVIKPSGIMVFNMGGRPGQGVCERLFRRR
At(  241) QGFVEDQFGLGLIARAVEEGISVIKPAGVMIFNMGGRPGQGVCRRLFERR
Wb(  250) QGFVEDQFGLGLIARAVEEGIDVIKPMGIMIFNMGGRPGQGVCKRLFERR

Zm(  300) GFRITKLWQTKIMQAADTDISALVEIEKNSRHRFEFFMDLVGDQPICART
At(  291) GVRVTQMWQTKILQAADTDISALVEIERSSPHRFEFFMGLSGDQPICART
Wb(  300) GLSVNKLWQTKILQASDTDISALVEIEKNNPHRFEFFMGLVGDRPICART

Zm(  350) AWAYMKSGGHISHALSVYSCQLRQPNQVKKIFEFLKDGFHEVSSSLDLSF
At(  341) AWAYGKAGGRISHALSVYSCQIRQPNLVKIIFDFLKNGFQEISNSLDLSF
Wb(  350) AWAFGKACGRISHALSVYSCQLRHPNEVKKIFEFLKNGFHDISNSLDLSF

Zm(  400) DDDSVAEEKIPFLAYLASFLKENKSNPCEPPAGCLNFRKLVAGFMKSYHH
At(  391) EDETVADEKIPFLAYLASVLKNSSYFPFEPPAGSKRFCSLIAGFMRTYHR
Wb(  400) EDDSVADEKIPFLAYLAGVLKDGSRFPYEPPTGNKRFRDLIASFMKTYHH

Zm(  450) IPLTPDNVVVFPSRSVAIENALQLFSPALAIVDEHLTRHLPKQWLTSLAI
At(  441) IPINQDNIVVFPSRAVAIESAFRLFSPRLAIVDEHLTRQLPRSWLTSLAI
Wb(  450) VPLSTDNVAIFPSRATAIENSLRLFTPRLAIVEEHLTCNLPRQWLTSLEI
```

FIG. 3A

```
Zm( 500) EGRADCNHADGIVTVIEAPRQSDLLIELIRKLQPQVVVTGMAQFEAITSA
At( 491) EDTSMDKSDD-QITVIESPHQSDLMIELIKKLKPQVVVTGMAPFEVITSS
Wb( 500) EQTRDSKTPIDGITVIEAPRQSDLMIELIKKLKPQVVVTGIAQFEAVTSS

Zm( 550) AFENLLNVTKDVGSRLFLDISEHLELSSLPSSNGVLKYLAGKTLPSHAAI
At( 540) SFLHLLEVTKEIGCRLFLDISDHFELSSLPASNGVLKYLAENQLPSHAAI
Wb( 550) AFEHLLRVTREIGSRLFIDISDQFELSSLPSSIGVLKYLARTPLPSHAAI

Zm( 600) LCGLVKNQVYSDLEVAFAISEDAAVYKALSQTIELLEGHTSLISQHYYGC
At( 590) ICGLVKNKVYSDLEVAFVIIEVDAIAKALSKTVEVLEGHTAIISQYYYGC
Wb( 600) ICGLLRNRVYIDLEVAFVISEEQTIFDALTRTVELLQGNTALISQYYYGC

Zm( 650) LFHELLAFQIADRHPQQERQPA--EVIPQQMIGFSDPAVSTLKATEFFVP
At( 640) LFHELLAFQLADRHAPAERESE--KAKSEEIIGFSSSAVSILKDAELSVT
Wb( 650) LFHELLSFQIPDRRQTAEREAENVEASDIDMIGFSSSAISVLSQSELSVR

Zm( 698) GSAESSIIHMDLDRSFLPVPSAVNASVFESFVRQNITDSETDVRSSIQQL
At( 688) EIDETSLIHMDVDQSFLQIPQSVKAAIFESFVRQNISEAEVDINPSIKQF
Wb( 700) VTEKSSLLHMDVDQIFLPTPIPVKAAIFESFARQNVTETECDVTPILRQF

Zm( 748) VKDSYGLSAAGCAEIIYGNTSVALFNKLVLCCMQEQGTLLFPLGTNGHYV
At( 738) VWSNYGFPTKSSTGFVYADGSLALFNKLVICCAQEGGTLCLPAGTNGNYV
Wb( 750) ILNTWNFSVEHSAEFIYADFPLALFNKLVLCCIEEGGSLCMPAGSNGNYA

Zm( 798) SAAKFVNASTVTIPTNPSSGFRIEPKVLADTLKNVSRPWVYVCGPTINPT
At( 788) AAAKFLKANVVNIPTESSDGFKLTERTLTKALESVKKPWVCISGPTVSPT
Wb( 800) AAAKFLNANIMSIPTEAEVGFKLTAKQLSSVLETVHKPWVYISGPTINPT

Zm( 848) GFLYSDSDIRELLSVCAEYGARVVIDTSFSGLEYETDGWRQWNLAGCLSS
At( 838) GLVYSNEEMDILLSTCAKFGAKVIIDTSFSGLEYSAT---SWDLKNALSK
Wb( 850) GLLYSNEEMKSLLIVCARYGARTIIDTSFSGIKFNSQDWDGWNLDASLAG

Zm( 898) LKRSEPSFSVVLLGELSFALTAGGHDFGFVILGDSSLAETFHS-FSSLSR
At( 885) LDSS---FSVSLLGCLSLNLLSGAIKLGFLVL-DQSLIDAFHT-LPGLSK
Wb( 900) LTGN-PSFSVCLLGGLFFKIPTGGLSYGFLVLKSGFLADSFRSSFSGLNK

Zm( 947) PHTTLKYTFKKLLGLKNQKDQHFSDLIVEQKEELKNRANQLIQTLESCGW
At( 930) PHSTVKYAAKKMLALKEEKASDFLDAVSETIKTLEGRSKRLKEVLQNSGW
Wb( 949) PHNTVRYTAKKLLELGEQKG-NLTGAAQGQEKLLATRLKRLKETLENCGW
```

FIG. 3B

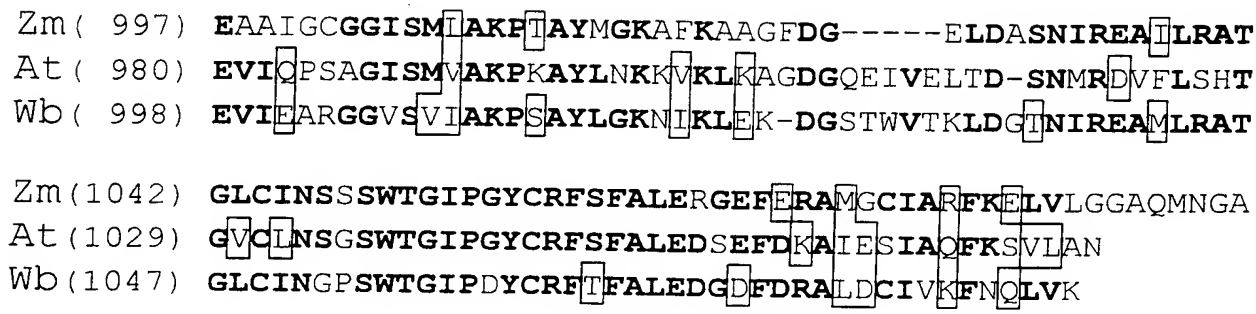
Zm(997) EAAIGCGGISMLAKPTAYMGKAFKAAGFDG-----ELDASNIREAILRAT
At(980) EVIQPSAGISMVAKPKAYLNKKVKLKAGDGQEIVELTD-SNMRDVFLSHT
Wb(998) EVIEARGGVSVIAKPSAYLGKNIKLEK-DGSTWVTKLDGINIREAMLRAT
Zm(1042) GLCINSSSWTGIPGYCRFSFALERGEFERAMGCIARFKELVLGGAQMNGA
At(1029) GVCLNSGSWTGIPGYCRFSFALEDSEFDKAIESIAQFKSVLAN
Wb(1047) GLCINGPSWTGIPDYCRFTFALEDGDFDRALDCIVKFNQLVK

FIG. 3C

Blast 2.0.9 [May-07-1999] Comparison Result

Match: Mismatch: gap open: gap extension:
x_dropoff: expect: wordsize: Filter

QUERY SEQUENCE SEQ ID NO:1 (length 3264)
SUBJECT SEQUENCE SEQ ID NO:5 (length 3273)

SHOW ONLY REGIONS HAVING ≥12 Contiguous Nucleotides Identical Sequence

Score = 577 bits (300), Expect = e-162 Identities = 762/996 (76%),
Positives = 762/996 (76%)

(SEQ IN NO:1) GAAAAAATTTACAATGATGGTAATACCAAGCATCTTTATGCCAGAGGACTGGTCCTTTAC 326
ATACCAAGCATCTT (SEQ ID NO:9)
(SEQ ID NO:5) GAAGAAGTTGACAATGATGGAGATACCAAGCATCTTCATTCCTGAAGATTGGTCATTCAC 326

(SEQ IN NO:1) ATTCTATGAGGGAATAAATAGACACCCAGACTCTATCTTCAAGGATAAGACAGTTGCTGA 386
(SEQ ID NO:10) GGATAAGACAGT
(SEQ ID NO:5) TTTCTATGAGGGCCTCAACCGTCATCCAGACTCCATTTTCAGGGATAAGACAGTAGCTGA 386

(SEQ IN NO:1) ACTTGGCTGTGGAAACGGATGGATATCCATAGCCATTGCTGAAAAGTGGTTGCCGTTGAA 446

(SEQ ID NO:5) GCTGGGATGTGGCAACGGTTGGATATCTATTGCTCTTGCAGAGAAGTGGTGCCCTTCAAA 446

FIG. 5A

(SEQ IN NO:1) GGTTTATGGCCTTGATATAAATCCAAGAGCAGTGAAGATTTCTTGGATAAATTTGTATTT 506

(SEQ ID NO:5) GGTCTATGGCCTGGATATAAACCCAAGAGCTGTGAAGATTGCATGGATAAACCTGTACTT 506

(SEQ IN NO:1) AAATGCTTTTGATGAGGATGGACAGCCCGTTTACGACAGCGAGAGCAAAACTTTGCTTGA 566

(SEQ ID NO:5) GAACGCATTAGATGACGACGGTCTCCCGATCTATGACGGGGAGGGGAAAACATTGCTCGA 566

(SEQ IN NO:1) CAGGGTAGAGTTTTATGAATCTGATCTGCTATCTTATTGCAGAGATAACCACATAGAGCT 626

TATGAATCTGATCT (SEQ ID NO:11)

(SEQ ID NO:5) TAGAGTTGAATTCTATGAATCTGATCTCCTTTCTTACTGTAGAGACAACAAGATAGAGCT 626

(SEQ IN NO:1) TGAACGAATTGTCGGATGCATCCCACAGATTCTTAATCCAAATCCAGATGCAATGTCCAA 686

(SEQ ID NO:5) CGATCGTATTGTTGGATGCATACCACAGATTCTCAACCCAAATCCAGAGGCGATGTCAAA 686

(SEQ IN NO:1) GCTTGTCACAGAAAAATGCTAGCGAAGAGTTTCTACATTCACTAAGCAACTATTGTGCCCT 746

(SEQ ID NO:5) GATTGTAACAGAGAATTCAAGTGAGGAGTTCTTGTACGCCTTGAGTAACTACTGTGCTCT 746

(SEQ IN NO:1) TCAGGGTTTTGTTGAGGATCAGTTTGGTTTAGGCCTAATTGCTAGGGCAGTTGAAGAAGG 806

TCAGGGTTTTGT (SEQ ID NO:12)

(SEQ ID NO:5) TCAGGGTTTTGTCGAGGACCAATTTGGCCTTGGGTTGATTGCTCGTGCGGTGGAAGAAGG 806

FIG. 5B

(SEQ IN NO:1) AATTGATGTGATCAAGCCTATGGGAATCATGATCTTCAACATGGGAGGGCGTCCAGGGCA 866
TTCAACATGGGAGG (SEQ ID NO:13)
(SEQ ID NO:5) GATATCTGTCATAAAGCCTTCAGGTATTATGGTATTCAACATGGGAGGTCGACCAGGACA 866

(SEQ IN NO:1) AGGTGTTTGTAAACGGTTATTTGAACGTCGTGGGCTCAGTGTTAACAAGCTGTGGCAGAC 926

(SEQ ID NO:5) GGGTGTCTGTGAACGTCTATTTCGACGGCGTGGGTTTCGCATTACTAAGCTCTGGCAAAC 926

(SEQ IN NO:1) AAAAATTCTTCAGGCTTCTGACACTGATATTTCAGCATTAGTTGAAATTGAGAAGAATAA 986

(SEQ ID NO:5) CAAAATTATGCAGGYTGCTGACACAGATATTTCAGCTTTGGTTGAAWTTGAGAAAAACAG 986

(SEQ IN NO:1) TCCACATCGTTTTGAGTTCTTCATGGGGCTTGTTGGAGACCGTCCAATTTGTGCTAGAAC 1046
GAGTTCTTCATGG (SEQ ID NO:14)
(SEQ ID NO:5) CAGACATCGATTCGAGTTCTTCATGGATCTTGTTGGGRATCAGCCTATCTGTGCTCGCAC 1046

(SEQ IN NO:1) TGCATGGGCGTTTGGAAAGGCCTGTGGGCGAATTTCTCATGCTTTGTCTGTTTACAGCTG 1106
CATGCTTTGTCTGT (SEQ ID NO:15)
(SEQ ID NO:5) AGCYTGGGCATACATGAAATCTGGTGGTCACATTTCACATGCTTTGTCTGTGTATAGCTG 1106

(SEQ IN NO:1) CCAGCTTCGGCATCCTAATGAGGTTAAGAAAATTTTTGAGTTTCTCAAAAATGGGTTTCA 1166

(SEQ ID NO:5) TCAACTTCGCCAGCCCAACCAGGTGAAGAAATATTTGAGTTTCTTAAAGATGGATTCCA 1166

FIG. 5C

(SEQ IN NO:1) CGACATCAGCAATTCTTTAGATTTATCCTTTGAAGATGATTCTGTTGCCGATGAGAAGAT 1226
GATTTATCCTTTGA (SEQ ID NO:16)
(SEQ ID NO:5) TGAAGTCAGCAGTTCCCTTGATTTATCCTTTGACGATGATTCTGTAGCTGAGGAAAAAAT 1226

(SEQ IN NO:1) TCCTTTCTTAGCATATCTTGCTGGTGTTTTGAAAGA 1262

(SEQ ID NO:5) TCCCTTCCTAGCTTACCTTGCTAGTTTTCTGAAAGA 1262

Score = 158 bits (82), Expect = 5e-36Identities = 326/448 (72%),
Positives = 326/448 (72%)

(SEQ IN NO:1) TTACTGTTATTGAAGCACCACGTCAGTCAGATTTGATGATAGAGCTGATAAAAAAGTTAA 1594

(SEQ ID NO:5) TTACTGTAATTGAGGCACCACGCCAATCAGATTTACTGATTGAGTTGATCAGGAAGCTGC 1594

(SEQ IN NO:1) AGCCGCAAGTGGTGGTCACAGGGATTGCTCAATTTGAGGCCGTTACTAGTTCAGCATTTG 1654
GCTCAATTTGAGGC (SEQ ID NO:17)
(SEQ ID NO:5) AGCCTCAGGTGGTTGTTACTGGCATGGCTCAATTTGAGGCTATCACCAGTGCTGCTTTTG 1654

(SEQ IN NO:1) AGCACCTTTTACGTGTCACCAGGGAAATTGGGTCTCGTCTTTTCATAGACATATCTGACC 1714

(SEQ ID NO:5) AGAACCTACTAAACGTAACAAAAGATGTTGGCTCCCGGCTGTTCCTGGATATTTCTGAGC 1714

FIG. 5D

(SEQ IN NO:1) AATTTGAGCTTTCAAGCCTCCCCAGTTCAATTGGGGTCTTGAAATATCTTGCTAGAACTC 1774

(SEQ ID NO:18) TTGAAATATCTTGCT

(SEQ ID NO:5) ATCTGGAGTTGTCTAGCCTGCCAAGCTCTAATGGTGTATTGAAATATCTTGCTGGAAAGA 1774

(SEQ IN NO:1) CACTGCCTTCTCATGCAGCTATAATATGTGGCTTGCTAAGAAATCGGGTATATACAGATC 1834

CATGCAGCTATA (SEQ ID NO:19)

(SEQ ID NO:5) CATTACCGTCACATGCAGCTATACTGTGTGGTTTAGTAAAGAATCAGGTGTACTCTGATC 1834

(SEQ IN NO:1) TTGAAGTAGCTTTTGTGATATCAGAAGAACAAACAATATTTGATGCATTGACAAGGACTG 1894

(SEQ ID NO:5) TGGAGGTTGCTTTTGCCATTTCCGAAGATGCAGCTGTATATAAAGCATTATCACAAACTA 1894

(SEQ IN NO:1) TAGAACTCTTGCAAGGCAATACTGCCCTGATTAGCCAGTATTACTATGGCTGTCTTTTCC 1954

(SEQ ID NO:5) TTGAGCTATTGGAAGGCCACACTTCTCTGATCAGCCAGCACTATTATGGTTGCCTTTTCC 1954

(SEQ IN NO:1) ATGAGCTTCTGTCCTTTCAGATTCCTGA 1982

(SEQ ID NO:5) ACGAGCTTCTGGCATTTCAGATTGCTGA 1982

FIG. 5E

FIG. 5F

Score = 52.6 bits (27), Expect = 3e-04Identities = 55/69 (79%),
Positives = 55/69 (79%)

(SEQ IN NO:1) CTTGCCATTGTTGAGGAACATCTGACCTGCAATCTACCCAGGCAATGGTTAACATCACTA 1491
(SEQ ID NO:20) GCAATGGTTAACATC
(SEQ ID NO:5) CTTGCAATTGTTGATGAACATTTAACCAGACACTTGCCCAAGCAATGGTTAACATCTTTA 1491

(SEQ IN NO:1) GAAATTGAG 1500

(SEQ ID NO:5) GCAATTGAG 1500

Blast 2.0.9 [May-07-1999] Comparison Result

Matrix 0 BLOSUM621 PAM302 PAM703 PAM2504 BLOSUM905 BLOSUM50 gap open: gap extension:
x_dropoff: expect: wordsize: Filter

QUERY SEQUENCE SEQ ID NO:2 (length 1088)
SUBJECT SEQUENCE SEQ ID NO:6 (length 1091)

SHOW ONLY REGIONS HAVING ≥8 Contiguous Amino Acid Identical Sequence

Score = 1390 bits (3559), Expect = 0.0 Identities = 679/1090 (62%), Positives = 851/1090 (77%), Gaps = 11/1090 (1%)

(SEQ ID NO: 2) MAAVTGLYGSIDEFLNHCSQSGDSAYSALRSLLERLEKPDTRTEARIFLAHLQKKLD-ND 59

(SEQ ID NO: 6) MAALAGEDKDVDAFLADCTASGDAAYGAAKAVLERLHAPATRPAARRLLGAVRRRFAASR 60

(SEQ ID NO: 2 GASQRCLETYHFQIQDIYLDRNEGTGYQNRKKFTMMVIPSIFMPEDWSFTFYEGINRHPD 119
(SEQ ID NO:26) PEDWSFTFYEG

(SEQ ID NO: 6) AAGEDCFRTFHFRIHDVVLDPHV-QGFQQMKKLTMMEIPSIFIPEDWSFTFYEGLNRHPD 119

FIG. 6A

(SEQ ID NO: 2) SIFKDKTVAELGCGNGWISIAIAEKWLPLKVYGLDINPRAVKISWINLYLNAFDEDGQPV 179
DKTVAELGCGNGWISIA KVYGLDINPRAVKI WINLYLNA
(SEQ ID NO:27) (SEQ ID NO:28) (SEQ ID NO:29)
(SEQ ID NO: 6) SIFRDKTVAELGCGNGWISIALAEKWCPSKVYGLDINPRAVKIAWINLYLNALDDDGLPI 179

(SEQ ID NO: 2) YDSESKTLLDRVEFYESDLLSYCRDNHIELERIVGCIPQILNPNPDAMSKLVTENASEEF 239
KTLLDRVEFYESDLLSYCRDN RIVGCIPQILNPNP
(SEQ ID NO:30) (SEQ ID NO:31)
(SEQ ID NO: 6) YDGEGKTLLDRVEFYESDLLSYCRDNKIELDRIVGCIPQILNPNPEAMSKIVTENSSEEF 239

(SEQ ID NO: 2) LHSLSNYCALQGFVEDQFGLGLIARAVEEGIDVIKPMGIMIFNMGGRPGQGVCKRLFERR 299
LSNYCALQGFVEDQFGLGLIARAVEEGI FNMGGRPGQGVC
(SEQ ID NO:32) (SEQ ID NO:33)
(SEQ ID NO: 6) LYALSNYCALQGFVEDQFGLGLIARAVEEGISVIKPSGIMVFNMGGRPGQGVCERLFRRR 299

(SEQ ID NO: 2) GLSVNKLWQTKILQASDTDISALVEIEKNNPHRFEFFMGLVGDRPICARTAWAFGKACGR 359
(SEQ ID NO:34) DTDISALVE (SEQ ID NO:35) PICARTAWA
(SEQ ID NO: 6) GFRITKLWQTKIMQXADTDISALVEXEKNSRHRFEFFMDLVGXQPICARTAWAYMKSGGH 359

(SEQ ID NO: 2) ISHALSVYSCQLRHPNEVKKIFEFLKNGFHDISNSLDLSFEDDSVADEKIPFLAYLAGVL 419
ISHALSVYSCQLR VKKIFEFLK EKIPFLAYLA
(SEQ ID NO:36) (SEQ ID NO:37) (SEQ ID NO:38)
(SEQ ID NO: 6) ISHALSVYSCQLRQPNQVKKIFEFLKDGFHEVSSSLDLSFDDDSVAEEKIPFLAYLASFL 419

(SEQ ID NO: 2) KDGSRFPYEPPTGNKRFRDLIASFMKTYHHVPLSTDNVAIFPSRATAIENSLRLFTPRLA 479

(SEQ ID NO: 6) KENKSNPCEPPAGCLNFRKLVAGFMKSYHHIPLTPDNVVVFPSRSVAIENALQLFSPALA 479

FIG. 6B

(SEQ ID NO: 2) IVEEHLTCNLPRQWLTSLEIEQTRDSKTPIDGITVIEAPRQSDLMIELIKKLKPQVVVTG 539
TVIEAPRQSDL (SEQ ID NO:39)
(SEQ ID NO: 6) IVDEHLTRHLPKQWLTSLAIEGRADCNHADGTVTVIEAPRQSDLLIELIRKLQPQVVVTG 539

(SEQ ID NO: 2) IAQFEAVTSSAFEHLLRVTREIGSRLFIDISDQFELSSLPSSIGVLKYLARTPLPSHAAI 599
ELSSLPSS (SEQ ID NO:40)
(SEQ ID NO: 6) MAQFEAITSAAFENLLNVTKDVGSRLFLDISEHLELSSLPSSNGVLKYLAGKTLPSHAAI 599

(SEQ ID NO: 2) ICGLLRNRVYTDLEVAFVISEEQTIFDALTRTVELLQGNTALISQYYYGCLFHELLSFQI 659
(SEQ ID NO:41) YYGCLFHELL
(SEQ ID NO: 6) LCGLVKNQVYSDLEVAFAISEDAAVYKALSQTIELLEGHTSLISQHYYGCLFHELLAFQI 659

(SEQ ID NO: 2) PDRRQTAEREAENVEASDIDMIGFSSSAISVLSQSELSVRVTEKSSLLHMDVDQIFLPTP 719

(SEQ ID NO: 6) ADRHPQQERQP--AEVIPQQMIGFSDPAVSTLKATEFFVPGSAESSIIHMDLDRSFLPVP 717

(SEQ ID NO: 2) TPVKAAIFESFARQNVTETECDVTPILRQFILNTWNFSVEHSAEFIYADFPLALFNKLVL 779
(SEQ ID NO:42) ALFNKLVL
(SEQ ID NO: 6) SAVNASVFESFVRQNITDSETDVRSSIQQLVKDSYGLSAAGCAEIIYGNTSVALFNKLVL 777

(SEQ ID NO: 2) CCIEEGGSLCMPAGSNGNYAAAAKFLNANIMSIPTEAEVGFKLTAKQLSSVLETVHKPWV 839

(SEQ ID NO: 6) CCMQEQGTLLFPLGTNGHYVSAAKFVNASTVTIPTNPSSGFRIEPKVLADTLKNVSRPWV 837

FIG. 6C

(SEQ ID NO: 2) YISGPTINPTGLLYSNEEMKSLLTVCARYGARTIIDTSFSGIKFNSQDWDGWNLDASLAG 899
GPTINPTG (SEQ ID NO:43)
(SEQ ID NO: 6) YVCGPTINPTGFLYSDSDIRELLSVCAEYGARVVIDTSFSGLEYETDGWRQWNLAGCLSS 897

(SEQ ID NO: 2) L-TGNPSFSVCLLGGLFFKIPTGGLSYGFLVLKSGFLADSFRSSFSGLNKPHNTVRYTAK 958

(SEQ ID NO: 6) LKRSEPSFSVVLLGELSFALTAGGHDFGFVILGDSSLAETFH-SFSSLSRPHTTLKYTFK 956

(SEQ ID NO: 2) KLLELGEQKG-NLTGAAQGQEKLLATRLKRLKETLENCGWEVIEARGGVSVIAKPSAYLG 1017

(SEQ ID NO: 6) KLLGLKNQKDQHFSDLIVEQKEELKNRANQLIQTLESCGWEAAIGCGGISMLAKPTAYMG 1016

(SEQ ID NO: 2) KNIKLEKDGSTWVTKLDGTNIREAMLRATGLCINGPSWTGIPDYCRFTFALEDGDFDRAL 1077
LRATGLCIN (SEQ ID NO:44)
(SEQ ID NO: 6) KAFK----AAGFDGELDASNIREAILRATGLCINSSSWTGIPGYCRFSFALERGEFERAM 1072

(SEQ ID NO: 2) DCIVKFNQLV 1087

(SEQ ID NO: 6) GCIARFKELV 1082

FIG. 6D

Blast 2.0.9 [May-07-1999] Comparison Result

Match: Mismatch: gap open: gap extension:
x_dropoff: expect: wordsize: Filter

QUERY SEQUENCE SEQ ID NO:7 (length 610)
SUBJECT SEQUENCE SEQ ID NO:5 (length 3273)

SHOW ONLY REGIONS HAVING ≥14 Contiguous Nucleotides Identical Sequence

Score = 358 bits (186), Expect = 5e-97Identities = 406/519 (78%),
Positives = 406/519 (78%)

(SEQ ID NO:7) CAGGGTTTTGTCGAGGATCAGTTTGGCTTGGGGCTTATTGCAAGGGCAGTTGAAGAAGGT 60
CAGGGTTTTGTCGAGGA (SEQ ID NO:21)
(SEQ ID NO:5) CAGGGTTTTGTCGAGGACCAATTTGGCCTTGGGTTGATTGCTCGTGCGGTGGAAGAAGGG 810

(SEQ ID NO:7) ATTTCTGTCATAAAGCCATTGGGCATTATGATCTTCAACATGGGAGGCCGTCCTGGGCAA 120
TCTGTCATAAAGCC (SEQ ID NO:22) TTCAACATGGGAGG (SEQ ID NO:23)
(SEQ ID NO:5) ATATCTGTCATAAAGCCTTCAGGTATTATGGTATTCAACATGGGAGGTCGACCAGGACAG 870

(SEQ ID NO:7) GGTGTTTGCAAACGGTTATTTGAGCGCCGTGGTCTTCGTGTTAACAAGCTCTGGCAAACT 180
(SEQ ID NO:24) AAGCTCTGGCAAAC
(SEQ ID NO:5) GGTGTCTGTGAACGTCTATTTCGACGGCGTGGGTTTCGCATTACTAAGCTCTGGCAAACC 930

FIG. 7A

(SEQ ID NO:7) AAAATTCTTCAGGCAGCTGACACTGATATATCAGCTCTAGTTGAAATTGAAAAGAGTAGC 240

(SEQ ID NO:5) AAAATTATGCAGGYTGCTGACACAGATATTTCAGCTTTGGTTGAAWTTGAGAAAAACAGC 990

(SEQ ID NO:7) ATGCACCGGTTTGAATTTTTTCATGGGACTTGTTGGAGATCAGCCAATATGTGCTCGAACA 300

(SEQ ID NO:5) AGACATCGATTCGAGTTCTTCATGGATCTTGTTGGGRATCAGCCTATCTGTGCTCGCACA 1050

(SEQ ID NO:7) GCATGGGCTTATGGCAAGGCTGGTGGTCGTATCTCTCATGCTTTATCTGTGTACAGCTGT 360

(SEQ ID NO:5) GCYTGGGCATACATGAAATCTGGTGGTCACATTTCACATGCTTTGTCTGTGTATAGCTGT 1110
(SEQ ID NO:25) AGCTGT

(SEQ ID NO:7) CAACTTCGTCAGCCAAATCAGGTCAGAAAGATATTTGAGTTCATAAAAAATGGATTCCAT 420
CAACTTCG

(SEQ ID NO:5) CAACTTCGCCAGCCCAACCAGGTGAAGAAAATATTTGAGTTTCTTAAAGATGGATTCCAT 1170

(SEQ ID NO:7) GATATCAGTAATTCTCTGGATTTGTCATTTGAGGATGATGCAGTAGCAGATGAAAAGATC 480

(SEQ ID NO:5) GAAGTCAGCAGTTCCCTTGATTTATCCTTTGACGATGATTCTGTAGCTGAGGAAAAAATT 1230

(SEQ ID NO:7) CCTTTCCTAGCTTATCTTGCTAGTGTGCTTAAGGAGAAC 519

(SEQ ID NO:5) CCCTTCCTAGCTTACCTTGCTAGTTTTCTGAAAGAGAAC 1269

FIG. 7B

Blast 2.0.9 [May-07-1999] Comparison Result

Matrix 0 BLOSUM621 PAM302 PAM703 PAM2504 BLOSUM905 BLOSUM50 gap open: gap extension:
x_dropoff: expect: wordsize: Filter

QUERY SEQUENCE SEQ ID NO:8 (length 203)
SUBJECT SEQUENCE SEQ ID NO:6 (length 1091)

SHOW ONLY REGIONS HAVING ≥11 Contiguous Amino Acid Identical Sequence

Score = 343 bits (871), Expect = 4e-94Identities = 164/203 (80%), Positives = 181/203 (88%)

```
(SEQ ID NO:8)  QGFVEDQFGLGLIARAVEEGISVIKPLGIMIFNMGGRPGQGVCKRLFERRGLRVNKLWQT 60
               QGFVEDQFGLGLIARAVEEGISVIKP     FNMGGRPGQGVC
                 (SEQ ID NO:45)                 (SEQ ID NO:46)
(SEQ ID NO:6)  QGFVEDQFGLGLIARAVEEGISVIKPSGIMVFNMGGRPGQGVCERLFRRRGFRITKLWQT 309


(SEQ ID NO:8)  KILQAADTDISALVEIEKSSMHRFEFFMGLVGDQPICARTAWAYGKAGGRISHALSVYSC 120
                    ADTDISALVE                   QPICARTAWAY      ISHALSVYSC
                     (SEQ ID NO:47)            (SEQ ID NO:48)     (SEQ ID NO:49)
(SEQ ID NO:6)  KIMQXADTDISALVEXEKNSRHRFEFFMDLVGXQPICARTAWAYMKSGGHISHALSVYSC 369
```

FIG. 8A

(SEQ ID NO:8) QLRQPNQVRKIFEFIKNGFHDISNSLDLSFEDDAVADEKIPFLAYLASVLKENSVFPYES 180
QLRQPNQV (SEQ ID NO:50) EKIPFLAYLAS (SEQ ID NO:51)
(SEQ ID NO:6) QLRQPNQVKKIFEFLKDGFHEVSSSLDLSFDDDSVAEEKIPFLAYLASFLKENKSNPCEP 429

(SEQ ID NO:8) PAGSRWFRNLIAGFMKTYHHFPL 203

(SEQ ID NO:6) PAGCLNFRKLVAGFMKSYHHIPL 452

FIG. 8B

SEQUENCE LISTING

<110> UNIVERSITY OF FLORIDA
MICHIGAN STATE UNIVERSITY
Hanson, Andrew D.
Gage, Douglas A.

<120> COMPOSTIONS AND METHODS FOR ALTERING SULFUR CONTENT IN PLANTS

<130> 4300.013610

<140> PCT/US00/04542
<141> 2000-02-22

<150> 60/121,038
<151> 1999-02-22

<160> 57

<170> PatentIn Ver. 2.1

<210> 1
<211> 3629
<212> DNA
<213> Wollastonia biflora

<400> 1
tgacagatct ctcgtcatcg ctgattgagt caacccacac acattatctt caacactctt 60
caccacaatc gattgccatc aatggctgca gttacaggct tgtacggatc aatagacgag 120
ttcttgaacc actgttcaca gtccggtgac tccgcctaca gtgcattaag atcactattg 180
gagcgcctcg agaaaccgga cacgcgaacc gaggcgagga tcttcttagc gcatctgcag 240
aaaaagctcg acaacgatgg agcctcgcaa cgttgcctcg agacttatca ttttcaaatc 300
caggatatat atctagatcg aaatgaaggt acaggctacc agaataggaa aaaatttaca 360
atgatggtaa taccaagcat ctttatgcca gaggactggt cctttacatt ctatgaggga 420
ataaatagac acccagactc tatcttcaag gataagacag ttgctgaact tggctgtgga 480
aacggatgga tatccatagc cattgctgaa aagtggttgc cgttgaaggt ttatggcctt 540
gatataaatc caagagcagt gaagatttct tggataaatt tgtatttaaa tgcttttgat 600
gaggatggac agcccgttta cgacagcgag agcaaaactt tgcttgacag ggtagagttt 660
tatgaatctg atctgctatc ttattgcaga gataaccaca tagagcttga acgaattgtc 720
ggatgcatcc cacagattct taatccaaat ccagatgcaa tgtccaagct tgtcacagaa 780
aatgctagcg aagagtttct acattcacta agcaactatt gtgccttca gggttttgtt 840
gaggatcagt ttggtttagg cctaattgct agggcagttg aagaaggaat tgatgtgatc 900
aagcctatgg gaatcatgat cttcaacatg ggagggcgtc cagggcaagg tgtttgtaaa 960
cggttatttg aacgtcgtgg gctcagtgtt aacaagctgt ggcagacaaa aattcttcag 1020
gcttctgaca ctgatatttc agcattagtt gaaattgaga agaataatcc acatcgtttt 1080
gagttcttca tggggcttgt tggagaccgt ccaatttgtg ctagaactgc atgggcgttt 1140
ggaaaggcct gtgggcgaat ttctcatgct ttgtctgttt acagctgcca gcttcggcat 1200
cctaatgagg ttaagaaaat ttttgagttt ctcaaaaatg ggtttcacga catcagcaat 1260
tctttagatt tatcctttga agatgattct gttgccgatg agaagattcc tttcttagca 1320
tatcttgctg gtgttttgaa agacggttct cgttttccct atgagcctcc aactggaaac 1380
aaaagatttc gtgatcttat tgcatccttc atgaaaacgt accaccatgt ccccccttagc 1440
actgataatg ttgctatttt tccttcaagg gctacggcta ttgagaactc cctccggttg 1500
ttcactccac gtcttgccat tgttgaggaa catctgacct gcaatctacc caggcaatgg 1560
ttaacatcac tagaaattga gcaaacaagg gacagcaaaa ctccaataga tggaattact 1620
gttattgaag caccacgtca gtcagatttg atgatagagc tgataaaaaa gttaaagccg 1680
caagtggtgg tcacagggat tgctcaattt gaggccgtta ctagttcagc atttgagcac 1740
cttttacgtg tcaccaggga aattgggtct cgtcttttca tagacatatc tgaccaattt 1800
gagctttcaa gcctccccag ttcaattggg gtcttgaaat atcttgctag aactccactg 1860
ccttctcatg cagctataat atgtggcttg ctaagaaatc gggtatatac agatcttgaa 1920
gtagcttttg tgatatcaga agaacaaaca atatttgatg cattgacaag gactgtagaa 1980
ctcttgcaag gcaatactgc cctgattagc cagtattact atggctgtct tttccatgag 2040

```
cttctgtcct ttcagattcc tgatagacgt caaactgcag agagagaagc tgagaatgtg 2100
gaagctagtg acattgacat gataggattt tcaagttcag caatctcggt tctgagccaa 2160
tcagaattat cagttagagt aaccgaaaaa tcctctttac ttcacatgga tgtggatcaa 2220
atatttcttc ccacaccaac cccggttaag gctgccatct ttgaaagctt tgcaaggcag 2280
aatgtgaccg aaacagagtg tgatgtcaca cccattctta ggcaattcat cctgaatact 2340
tggaactttt cggttgagca cagtgcagaa ttcatatacg cagacttccc tcttgctctt 2400
ttcaataaac tcgtcctctg ctgcattgaa gaaggggaa gcctatgcat gccagctgga 2460
tcaaatggta attatgcagc agctgccaaa ttcttaaatg caaacatcat gtctatccct 2520
acagaggctg aagtaggttt taaactgacg gctaagcaac ttagtagtgt attggagact 2580
gtacataaac catgggtgta tatttctggt ccaacaatta acccgacagg tttgctttac 2640
agcaatgaag agatgaaaag cttgctgaca gtatgtgcta gatacggggc aaggactata 2700
attgatacat cgttctcggg gatcaagttt aactcccagg attgggatgg gtggaatttg 2760
gatgcaagtt tagcaggatt aactggaaat ccgtctttta gcgtgtgttt gcttggagga 2820
ttgtttttta agatacctac tggagggctt tcatatgggt ttcttgttct gaaatcagga 2880
tttttggctg attcatttcg tagtagtttc tcggggttga ataaacctca caatactgtt 2940
aggtatactg caaagaaatt gcttgaactt ggagagcaga aagggaattt aacaggtgct 3000
gctcaaggac aggagaaatt gctggctact agactgaagc gtctgaaaga gacacttgag 3060
aactgtggat gggaagtaat tgaagctcgt ggaggcgttt ctgtgattgc aaagccgtct 3120
gcttatcttg gcaaaaacat taagctcgag aaagacggtt ctacatgggt aaccaagctc 3180
gatggcacca acatccgaga agcaatgctg agagctacag gtttgtgcat caatggacct 3240
tcttggactg ggattccaga ctattgtcgc ttcacttttg cccttgaaga tggcgatttt 3300
gaccgagctt tggattgcat cgtgaaattc aaccaacttg ttaagtaacg gtcacttgtt 3360
ttgtttttgt gacttgggtg ttgcaaaata aaccgcttat gaccatgtca tggagctatg 3420
gttatttatc ttctatttga gaaagaagtt acataggtga aagtctaagc agaaataaat 3480
gttggatcta tgtgttgaag acaacttgaa acagtgttgc tattgacatg tgatcttatt 3540
ggattacata aggagcataa ttatcaaaaa aatgtttaaa tatgatttaa atgtaatgaa 3600
tcgaatataa ttacttgtgt tattaatta                                   3629
```

<210> 2
<211> 1088
<212> PRT
<213> Wollastonia biflora

<400> 2

```
Met Ala Ala Val Thr Gly Leu Tyr Gly Ser Ile Asp Glu Phe Leu Asn
  1               5                  10                  15

His Cys Ser Gln Ser Gly Asp Ser Ala Tyr Ser Ala Leu Arg Ser Leu
             20                  25                  30

Leu Glu Arg Leu Glu Lys Pro Asp Thr Arg Thr Glu Ala Arg Ile Phe
         35                  40                  45

Leu Ala His Leu Gln Lys Lys Leu Asp Asn Asp Gly Ala Ser Gln Arg
     50                  55                  60

Cys Leu Glu Thr Tyr His Phe Gln Ile Gln Asp Ile Tyr Leu Asp Arg
 65                  70                  75                  80

Asn Glu Gly Thr Gly Tyr Gln Asn Arg Lys Lys Phe Thr Met Met Val
                 85                  90                  95

Ile Pro Ser Ile Phe Met Pro Glu Asp Trp Ser Phe Thr Phe Tyr Glu
            100                 105                 110

Gly Ile Asn Arg His Pro Asp Ser Ile Phe Lys Asp Lys Thr Val Ala
        115                 120                 125

Glu Leu Gly Cys Gly Asn Gly Trp Ile Ser Ile Ala Ile Ala Glu Lys
    130                 135                 140
```

```
Trp Leu Pro Leu Lys Val Tyr Gly Leu Asp Ile Asn Pro Arg Ala Val
145                 150                 155                 160

Lys Ile Ser Trp Ile Asn Leu Tyr Leu Asn Ala Phe Asp Glu Asp Gly
                165                 170                 175

Gln Pro Val Tyr Asp Ser Glu Ser Lys Thr Leu Leu Asp Arg Val Glu
            180                 185                 190

Phe Tyr Glu Ser Asp Leu Leu Ser Tyr Cys Arg Asp Asn His Ile Glu
        195                 200                 205

Leu Glu Arg Ile Val Gly Cys Ile Pro Gln Ile Leu Asn Pro Asn Pro
    210                 215                 220

Asp Ala Met Ser Lys Leu Val Thr Glu Asn Ala Ser Glu Glu Phe Leu
225                 230                 235                 240

His Ser Leu Ser Asn Tyr Cys Ala Leu Gln Gly Phe Val Glu Asp Gln
                245                 250                 255

Phe Gly Leu Gly Leu Ile Ala Arg Ala Val Glu Glu Gly Ile Asp Val
            260                 265                 270

Ile Lys Pro Met Gly Ile Met Ile Phe Asn Met Gly Gly Arg Pro Gly
        275                 280                 285

Gln Gly Val Cys Lys Arg Leu Phe Glu Arg Arg Gly Leu Ser Val Asn
    290                 295                 300

Lys Leu Trp Gln Thr Lys Ile Leu Gln Ala Ser Asp Thr Asp Ile Ser
305                 310                 315                 320

Ala Leu Val Glu Ile Glu Lys Asn Asn Pro His Arg Phe Glu Phe Phe
                325                 330                 335

Met Gly Leu Val Gly Asp Arg Pro Ile Cys Ala Arg Thr Ala Trp Ala
            340                 345                 350

Phe Gly Lys Ala Cys Gly Arg Ile Ser His Ala Leu Ser Val Tyr Ser
        355                 360                 365

Cys Gln Leu Arg His Pro Asn Glu Val Lys Lys Ile Phe Glu Phe Leu
    370                 375                 380

Lys Asn Gly Phe His Asp Ile Ser Asn Ser Leu Asp Leu Ser Phe Glu
385                 390                 395                 400

Asp Asp Ser Val Ala Asp Glu Lys Ile Pro Phe Leu Ala Tyr Leu Ala
                405                 410                 415

Gly Val Leu Lys Asp Gly Ser Arg Phe Pro Tyr Glu Pro Pro Thr Gly
            420                 425                 430

Asn Lys Arg Phe Arg Asp Leu Ile Ala Ser Phe Met Lys Thr Tyr His
        435                 440                 445

His Val Pro Leu Ser Thr Asp Asn Val Ala Ile Phe Pro Ser Arg Ala
    450                 455                 460

Thr Ala Ile Glu Asn Ser Leu Arg Leu Phe Thr Pro Arg Leu Ala Ile
465                 470                 475                 480
```

Val Glu Glu His Leu Thr Cys Asn Leu Pro Arg Gln Trp Leu Thr Ser
485 490 495

Leu Glu Ile Glu Gln Thr Arg Asp Ser Lys Thr Pro Ile Asp Gly Ile
500 505 510

Thr Val Ile Glu Ala Pro Arg Gln Ser Asp Leu Met Ile Glu Leu Ile
515 520 525

Lys Lys Leu Lys Pro Gln Val Val Val Thr Gly Ile Ala Gln Phe Glu
530 535 540

Ala Val Thr Ser Ser Ala Phe Glu His Leu Leu Arg Val Thr Arg Glu
545 550 555 560

Ile Gly Ser Arg Leu Phe Ile Asp Ile Ser Asp Gln Phe Glu Leu Ser
565 570 575

Ser Leu Pro Ser Ser Ile Gly Val Leu Lys Tyr Leu Ala Arg Thr Pro
580 585 590

Leu Pro Ser His Ala Ala Ile Ile Cys Gly Leu Leu Arg Asn Arg Val
595 600 605

Tyr Thr Asp Leu Glu Val Ala Phe Val Ile Ser Glu Glu Gln Thr Ile
610 615 620

Phe Asp Ala Leu Thr Arg Thr Val Glu Leu Leu Gln Gly Asn Thr Ala
625 630 635 640

Leu Ile Ser Gln Tyr Tyr Tyr Gly Cys Leu Phe His Glu Leu Leu Ser
645 650 655

Phe Gln Ile Pro Asp Arg Arg Gln Thr Ala Glu Arg Glu Ala Glu Asn
660 665 670

Val Glu Ala Ser Asp Ile Asp Met Ile Gly Phe Ser Ser Ser Ala Ile
675 680 685

Ser Val Leu Ser Gln Ser Glu Leu Ser Val Arg Val Thr Glu Lys Ser
690 695 700

Ser Leu Leu His Met Asp Val Asp Gln Ile Phe Leu Pro Thr Pro Thr
705 710 715 720

Pro Val Lys Ala Ala Ile Phe Glu Ser Phe Ala Arg Gln Asn Val Thr
725 730 735

Glu Thr Glu Cys Asp Val Thr Pro Ile Leu Arg Gln Phe Ile Leu Asn
740 745 750

Thr Trp Asn Phe Ser Val Glu His Ser Ala Glu Phe Ile Tyr Ala Asp
755 760 765

Phe Pro Leu Ala Leu Phe Asn Lys Leu Val Leu Cys Cys Ile Glu Glu
770 775 780

Gly Gly Ser Leu Cys Met Pro Ala Gly Ser Asn Gly Asn Tyr Ala Ala
785 790 795 800

Ala Ala Lys Phe Leu Asn Ala Asn Ile Met Ser Ile Pro Thr Glu Ala

805 810 815

Glu Val Gly Phe Lys Leu Thr Ala Lys Gln Leu Ser Ser Val Leu Glu
820 825 830

Thr Val His Lys Pro Trp Val Tyr Ile Ser Gly Pro Thr Ile Asn Pro
835 840 845

Thr Gly Leu Leu Tyr Ser Asn Glu Glu Met Lys Ser Leu Leu Thr Val
850 855 860

Cys Ala Arg Tyr Gly Ala Arg Thr Ile Ile Asp Thr Ser Phe Ser Gly
865 870 875 880

Ile Lys Phe Asn Ser Gln Asp Trp Asp Gly Trp Asn Leu Asp Ala Ser
885 890 895

Leu Ala Gly Leu Thr Gly Asn Pro Ser Phe Ser Val Cys Leu Leu Gly
900 905 910

Gly Leu Phe Phe Lys Ile Pro Thr Gly Gly Leu Ser Tyr Gly Phe Leu
915 920 925

Val Leu Lys Ser Gly Phe Leu Ala Asp Ser Phe Arg Ser Ser Phe Ser
930 935 940

Gly Leu Asn Lys Pro His Asn Thr Val Arg Tyr Thr Ala Lys Lys Leu
945 950 955 960

Leu Glu Leu Gly Glu Gln Lys Gly Asn Leu Thr Gly Ala Ala Gln Gly
965 970 975

Gln Glu Lys Leu Leu Ala Thr Arg Leu Lys Arg Leu Lys Glu Thr Leu
980 985 990

Glu Asn Cys Gly Trp Glu Val Ile Glu Ala Arg Gly Gly Val Ser Val
995 1000 1005

Ile Ala Lys Pro Ser Ala Tyr Leu Gly Lys Asn Ile Lys Leu Glu Lys
1010 1015 1020

Asp Gly Ser Thr Trp Val Thr Lys Leu Asp Gly Thr Asn Ile Arg Glu
1025 1030 1035 1040

Ala Met Leu Arg Ala Thr Gly Leu Cys Ile Asn Gly Pro Ser Trp Thr
1045 1050 1055

Gly Ile Pro Asp Tyr Cys Arg Phe Thr Phe Ala Leu Glu Asp Gly Asp
1060 1065 1070

Phe Asp Arg Ala Leu Asp Cys Ile Val Lys Phe Asn Gln Leu Val Lys
1075 1080 1085

<210> 3
<211> 3542
<212> DNA
<213> Arabidopsis thaliana

<400> 3
tcggcacgag cacactttca ctttcattcg ctttttcttc tctattactg caatcacttg 60
ttcacgataa tggcggacct ctcctctgtc gatgagttcc tgaatcagtg caaacaatcc 120
ggcgacgcag cttatggcgc cttgcgatcg gtcttggagc ggctcgagga tccgaatact 180
cgatctaagg ctcggatctt cttgtctgac atctacaaac gcgtcggatc ttctgagacg 240
tctcttccaaa cctaccattt tcatatccag gacatctatc tcgatcaata cgaaggcttt 300
cagtccagga aaaagttaac catgatggtc attcctagta tttttattcc agaagactgg 360
tcatttacat tttatgaagg acttaataga catcctgaca ccatctttaa ggataagact 420
gtttctgaac ttggctgtgg gaatggatgg atatcaatag ccattgctgc taagtggttg 480
ccttcaaagg tatatgggct tgatattaat cctagagctg tgaaaatttc ttggataaat 540
ttgtacctaa acgctcttga tgataatggc gaaccagtct atgacgaaga gaagaaaact 600
ttattggaca gagtggaatt ctatgaatct gatttgcttg gttattgtag agataataaa 660
attcagttag agagaatcgt aggatgcata cctcagattc ttaatccaaa cccagaagct 720
atgtctaagc tgatcacaga aaatgcaagt gaggaatttc tccattcgct gagtaactat 780
tgtgcccttc agggttttgt tgaagatcag tttggcttag gttttgattgc cagagcggtt 840
gaagaaggaa tatctgtcat caaacctgca ggggttatga tattcaacat gggtggtcgt 900
cctgggcaag gtgtctgtag acgcttgttt gagcggcgag gagtccgtgt tacacagatg 960
tggcagacta aaatacttca ggctgcagat actgatatct cagcattagt tgaaattgag 1020
aggagcagcc ctcatcgttt tgagttcttt atgggacttt ctggagacca accaatttgt 1080
gctcgaacag catgggccta tgggaaggct ggtggccgaa tctcccatgc tttatcggtt 1140
tatagttgtc agattcgcca accaaatctg gttaagataa tctttgactt cttgaaaaat 1200
ggattccaag aaatcagtaa ttctctggat ttatctttcg aagatgaaac tgttgctgat 1260
gagaagattc cattcctagc ctatcttgct agtgtcttga aaaatagctc ctatttcccg 1320
tttgaacctc cagctggcag caaaagattc tgcagtctaa ttgcaggctt tatgaggaca 1380
taccaccgta taccaattaa tcaggataac attgtcgtgt ttccatcaag ggctgtggca 1440
atcgagagtg catttcgttt attctctcct cgacttgcaa ttgttgatga gcatttaact 1500
cgacaacttc cgaggagctg gctaacctct ttagccattg aggacactag catggataaa 1560
tcagatgatc aaatcacagt catcgaatcc ccgcaccaat ctgatctgat gatagaactc 1620
attaagaaac taaagccaca ggtggtggtt actggaatgg caccatttga ggtcattacc 1680
agttcatcgt ttttgcacct gttggaagtg acaaaagaaa ttggatgtcg acttttcttg 1740
gatatatctg atcactttga gttgtctagc ctccctgcat ccaatggggt actaaaatat 1800
ctcgctgaaa atcaactacc ttctcatgca gcaattatct gcggtctggt aaagaataag 1860
gtctattcag atttagaagt agcctttgtc attacagaag tggatgccat tgctaaagcc 1920
ctatccaaaa ctgtggaagt cttagaaggt catactgcta ttatcagtca atactactat 1980
ggttgccttt tccatgagct tttggctttc cagcttgctg atcgccatgc cccggctgag 2040
agggaaagtg agaaggcaaa gtctgaagag atcattggat tttcaagctc agccgtctct 2100
attctaaaag atgctgagct ttcagtcact gagattgacg aaacttctct aatccacatg 2160
gatgttgatc aaagtttctt gcaaattcca caatctgtta aggctgcaat ctttgaaagt 2220
tttgtgaggc agaacatatc tgaagcggaa gttgatatca acccaagtat taagcagttt 2280
gtgtggagta attatgggtt tccaaccaaa agcagcacag gctttgtata tgctgatggc 2340
tcactagcac tattcaataa actggtaatt tgttgtgctc aagaaggtgg aacgctttgt 2400
ttacctgctg gcacaaatgg aaattatgtt gctgctgcca aatttttaaa agctaatgtc 2460
gtgaatatcc ctaccgagtc tagtgatggc tttaagctga cagaaaggac tctaacgaaa 2520
gcactcgagt ctgtgaagaa gccgtgggtt tgtatatctg gaccaacggt tagccctaca 2580
ggcttggtgt acagcaatga ggagatggat atactgttgt ctacttgtgc taagtttgga 2640
gcaaaggtca tcatcgatac ttcattctcg ggattagaat acagtgcaac tagctgggat 2700
ttgaagaacg ctttgtcgaa attggattct tccttttcag tttcgctgct tggatgtctc 2760
tctttgaatt tgctcagtgg agctattaaa ctcgggtttc tagttttgga tcagtctctc 2820
atcgatgcct tccataccct cccaggcctg agcaaaccgc acagcactgt gaaatatgcc 2880
gctaagaaaa tgttggcttt gaaggaagag aaagcaagtg actttctgga tgccgtttct 2940
gaaaccatta aaaccttgga aggcagatct aaacgcttaa aagaggtact acagaactct 3000
ggttgggagg ttatccaacc ctcagctgga atctcaatgg tggcaaagcc gaaagcttat 3060
ctcaacaaaa aagtaaagct gaaagcagga gatggtcagg aaatcgtcga gcttacggat 3120
tcaaatatga gggatgtgtt cctcagccat accggtgttt gcttaaacag cggttcctgg 3180
actggaatcc ctggttactg ccggttttca tttgcattgg aagatagtga gtttgacaag 3240
gctattgaat cgatagctca gtttaaaagc gtccttgcta actgaaacgt gccagtttca 3300
tttacgatgt catcaacttg aactaaaggg cattgtgcat tgctggttct gagacttaac 3360
acgtcgctct tcgtctgcgt tcctttgttc tgttgtttct aaataactcg cttcccgttt 3420
cttggtcagc cccataacgg gtttactttt tggaagtatg atattttcat gctaataaag 3480
gagtcataaa actttgattt tgtttgtctt atattttaat gtgaataagt ctctttgttt 3540
ca 3542

<210> 4
<211> 1071
<212> PRT
<213> Arabidopsis thaliana

<400> 4
Met Ala Asp Leu Ser Ser Val Asp Glu Phe Leu Asn Gln Cys Lys Gln
1 5 10 15

Ser Gly Asp Ala Ala Tyr Gly Ala Leu Arg Ser Val Leu Glu Arg Leu
20 25 30

Glu Asp Pro Asn Thr Arg Ser Lys Ala Arg Ile Phe Leu Ser Asp Ile
35 40 45

Tyr Lys Arg Val Gly Ser Ser Glu Thr Ser Leu Gln Thr Tyr His Phe
50 55 60

His Ile Gln Asp Ile Tyr Leu Asp Gln Tyr Glu Gly Phe Gln Ser Arg
65 70 75 80

Lys Lys Leu Thr Met Met Val Ile Pro Ser Ile Phe Ile Pro Glu Asp
85 90 95

Trp Ser Phe Thr Phe Tyr Glu Gly Leu Asn Arg His Pro Asp Thr Ile
100 105 110

Phe Lys Asp Lys Thr Val Ser Glu Leu Gly Cys Gly Asn Gly Trp Ile
115 120 125

Ser Ile Ala Ile Ala Ala Lys Trp Leu Pro Ser Lys Val Tyr Gly Leu
130 135 140

Asp Ile Asn Pro Arg Ala Val Lys Ile Ser Trp Ile Asn Leu Tyr Leu
145 150 155 160

Asn Ala Leu Asp Asp Asn Gly Glu Pro Val Tyr Asp Glu Glu Lys Lys
165 170 175

Thr Leu Leu Asp Arg Val Glu Phe Tyr Glu Ser Asp Leu Leu Gly Tyr
180 185 190

Cys Arg Asp Asn Lys Ile Gln Leu Glu Arg Ile Val Gly Cys Ile Pro
195 200 205

Gln Ile Leu Asn Pro Asn Pro Glu Ala Met Ser Lys Leu Ile Thr Glu
210 215 220

Asn Ala Ser Glu Glu Phe Leu His Ser Leu Ser Asn Tyr Cys Ala Leu
225 230 235 240

Gln Gly Phe Val Glu Asp Gln Phe Gly Leu Gly Leu Ile Ala Arg Ala
245 250 255

Val Glu Glu Gly Ile Ser Val Ile Lys Pro Ala Gly Val Met Ile Phe
260 265 270

Asn Met Gly Gly Arg Pro Gly Gln Gly Val Cys Arg Arg Leu Phe Glu
275 280 285

Arg Arg Gly Val Arg Val Thr Gln Met Trp Gln Thr Lys Ile Leu Gln

290 295 300

Ala Ala Asp Thr Asp Ile Ser Ala Leu Val Glu Ile Glu Arg Ser Ser
305 310 315 320

Pro His Arg Phe Glu Phe Phe Met Gly Leu Ser Gly Asp Gln Pro Ile
325 330 335

Cys Ala Arg Thr Ala Trp Ala Tyr Gly Lys Ala Gly Gly Arg Ile Ser
340 345 350

His Ala Leu Ser Val Tyr Ser Cys Gln Ile Arg Gln Pro Asn Leu Val
355 360 365

Lys Ile Ile Phe Asp Phe Leu Lys Asn Gly Phe Gln Glu Ile Ser Asn
370 375 380

Ser Leu Asp Leu Ser Phe Glu Asp Glu Thr Val Ala Asp Glu Lys Ile
385 390 395 400

Pro Phe Leu Ala Tyr Leu Ala Ser Val Leu Lys Asn Ser Ser Tyr Phe
405 410 415

Pro Phe Glu Pro Pro Ala Gly Ser Lys Arg Phe Cys Ser Leu Ile Ala
420 425 430

Gly Phe Met Arg Thr Tyr His Arg Ile Pro Ile Asn Gln Asp Asn Ile
435 440 445

Val Val Phe Pro Ser Arg Ala Val Ala Ile Glu Ser Ala Phe Arg Leu
450 455 460

Phe Ser Pro Arg Leu Ala Ile Val Asp Glu His Leu Thr Arg Gln Leu
465 470 475 480

Pro Arg Ser Trp Leu Thr Ser Leu Ala Ile Glu Asp Thr Ser Met Asp
485 490 495

Lys Ser Asp Asp Gln Ile Thr Val Ile Glu Ser Pro His Gln Ser Asp
500 505 510

Leu Met Ile Glu Leu Ile Lys Lys Leu Lys Pro Gln Val Val Val Thr
515 520 525

Gly Met Ala Pro Phe Glu Val Ile Thr Ser Ser Ser Phe Leu His Leu
530 535 540

Leu Glu Val Thr Lys Glu Ile Gly Cys Arg Leu Phe Leu Asp Ile Ser
545 550 555 560

Asp His Phe Glu Leu Ser Ser Leu Pro Ala Ser Asn Gly Val Leu Lys
565 570 575

Tyr Leu Ala Glu Asn Gln Leu Pro Ser His Ala Ala Ile Ile Cys Gly
580 585 590

Leu Val Lys Asn Lys Val Tyr Ser Asp Leu Glu Val Ala Phe Val Ile
595 600 605

Thr Glu Val Asp Ala Ile Ala Lys Ala Leu Ser Lys Thr Val Glu Val
610 615 620

Leu Glu Gly His Thr Ala Ile Ile Ser Gln Tyr Tyr Tyr Gly Cys Leu
625 630 635 640

Phe His Glu Leu Leu Ala Phe Gln Leu Ala Asp Arg His Ala Pro Ala
645 650 655

Glu Arg Glu Ser Glu Lys Ala Lys Ser Glu Glu Ile Ile Gly Phe Ser
660 665 670

Ser Ser Ala Val Ser Ile Leu Lys Asp Ala Glu Leu Ser Val Thr Glu
675 680 685

Ile Asp Glu Thr Ser Leu Ile His Met Asp Val Asp Gln Ser Phe Leu
690 695 700

Gln Ile Pro Gln Ser Val Lys Ala Ala Ile Phe Glu Ser Phe Val Arg
705 710 715 720

Gln Asn Ile Ser Glu Ala Glu Val Asp Ile Asn Pro Ser Ile Lys Gln
725 730 735

Phe Val Trp Ser Asn Tyr Gly Phe Pro Thr Lys Ser Ser Thr Gly Phe
740 745 750

Val Tyr Ala Asp Gly Ser Leu Ala Leu Phe Asn Lys Leu Val Ile Cys
755 760 765

Cys Ala Gln Glu Gly Gly Thr Leu Cys Leu Pro Ala Gly Thr Asn Gly
770 775 780

Asn Tyr Val Ala Ala Ala Lys Phe Leu Lys Ala Asn Val Val Asn Ile
785 790 795 800

Pro Thr Glu Ser Ser Asp Gly Phe Lys Leu Thr Glu Arg Thr Leu Thr
805 810 815

Lys Ala Leu Glu Ser Val Lys Lys Pro Trp Val Cys Ile Ser Gly Pro
820 825 830

Thr Val Ser Pro Thr Gly Leu Val Tyr Ser Asn Glu Glu Met Asp Ile
835 840 845

Leu Leu Ser Thr Cys Ala Lys Phe Gly Ala Lys Val Ile Ile Asp Thr
850 855 860

Ser Phe Ser Gly Leu Glu Tyr Ser Ala Thr Ser Trp Asp Leu Lys Asn
865 870 875 880

Ala Leu Ser Lys Leu Asp Ser Ser Phe Ser Val Ser Leu Leu Gly Cys
885 890 895

Leu Ser Leu Asn Leu Leu Ser Gly Ala Ile Lys Leu Gly Phe Leu Val
900 905 910

Leu Asp Gln Ser Leu Ile Asp Ala Phe His Thr Leu Pro Gly Leu Ser
915 920 925

Lys Pro His Ser Thr Val Lys Tyr Ala Ala Lys Lys Met Leu Ala Leu
930 935 940

Lys Glu Glu Lys Ala Ser Asp Phe Leu Asp Ala Val Ser Glu Thr Ile
945 950 955 960

```
Lys Thr Leu Glu Gly Arg Ser Lys Arg Leu Lys Glu Val Leu Gln Asn
                965                 970                 975

Ser Gly Trp Glu Val Ile Gln Pro Ser Ala Gly Ile Ser Met Val Ala
            980                 985                 990

Lys Pro Lys Ala Tyr Leu Asn Lys Lys Val Lys Leu Lys Ala Gly Asp
        995                 1000                1005

Gly Gln Glu Ile Val Glu Leu Thr Asp Ser Asn Met Arg Asp Val Phe
   1010                1015                1020

Leu Ser His Thr Gly Val Cys Leu Asn Ser Gly Ser Trp Thr Gly Ile
1025                1030                1035                1040

Pro Gly Tyr Cys Arg Phe Ser Phe Ala Leu Glu Asp Ser Glu Phe Asp
                1045                1050                1055

Lys Ala Ile Glu Ser Ile Ala Gln Phe Lys Ser Val Leu Ala Asn
            1060                1065                1070
```

<210> 5
<211> 3499
<212> DNA
<213> Zea mays

<400> 5
gatatggcgg cgctggcggg cgaggacaag gacgttgacg ccttcctcgc ggactgcacg 60
gcgtcgggag acgccgcgta tggcgccgcc aaggccgtgc tcgagcgcct ccacgccccc 120
gccacgcgcc ccgccgcacg acgcctcctt ggcgccgtcc gccgacgctt cgccgcctcc 180
cgcgccgcgg gggaggactg cttccgcacc ttccactttc gcatccacga cgtcgtcctc 240
gacccacacg tacaaggttt ccaacaaatg aagaagttga caatgatgga gataccaagc 300
atcttcattc ctgaagattg gtcattcact ttctatgagg gcctcaaccg tcatccagac 360
tccattttca gggataagac agtagctgag ctgggatgtg gcaacggttg gatatctatt 420
gctcttgcag agaagtggtg cccttcaaag gtctatggcc tggatataaa cccaagagct 480
gtgaagattg catggataaa cctgtacttg aacgcattag atgacgacgg tctcccgatc 540
tatgacgggg aggggaaaac attgctcgat agagttgaat tctatgaatc tgatctcctt 600
tcttactgta gagacaacaa gatagagctc gatcgtattg ttggatgcat accacagatt 660
ctcaacccaa atccagaggc gatgtcaaag attgtaacag agaattcaag tgaggagttc 720
ttgtacgcct tgagtaacta ctgtgctctt cagggttttg tcgaggacca atttggcctt 780
gggttgattg ctcgtgcggt ggaagaaggg atatctgtca taaagccttc aggtattatg 840
gtattcaaca tgggaggtcg accaggacag ggtgtctgtg aacgtctatt tcgacggcgt 900
gggtttcgca ttactaagct ctggcaaacc aaaattatgc aggytgctga cacagatatt 960
tcagctttgg ttgaawttga gaaaaacagc agacatcgat tcgagttctt catggatctt 1020
gttgggratc agcctatctg tgctcgcaca gcytgggcat acatgaaatc tggtggtcac 1080
atttcacatg ctttgtctgt gtatagctgt caacttcgcc agcccaacca ggtgaagaaa 1140
atatttgagt ttcttaaaga tggattccat gaagtcagca gttcccttga tttatccttt 1200
gacgatgatt ctgtagctga ggaaaaaatt cccttcctag cttaccttgc tagttttctg 1260
aaagagaaca agtcaaatcc ctgtgagcca ccagctggat gtctaaactt taggaaactt 1320
gttgctggat ttatgaagag ctaccatcac attcctctaa ctcctgataa tgtcgttgtg 1380
ttccccttctc gctctgtggc aatagagaat gctcttcaat tgttctcacc ggcgcttgca 1440
attgttgatg aacatttaac cagacacttg cccaagcaat ggttaacatc tttagcaatt 1500
gagggaagag cagattgtaa tcacgccgat ggtacagtta ctgtaattga ggcaccacgc 1560
caatcagatt tactgattga gttgatcagg aagctgcagc ctcaggtggt tgttactggc 1620
atggctcaat ttgaggctat caccagtgct gcttttgaga acctactaaa cgtaacaaaa 1680
gatgttggct cccggctgtt cctggatatt tctgagcatc tggagttgtc tagcctgcca 1740
agctctaatg gtgtattgaa atatcttgct ggaaagacat taccgtcaca tgcagctata 1800
ctgtgtggtt tagtaaagaa tcaggtgtac tctgatctgg aggttgcttt tgccatttcc 1860
gaagatgcag ctgtatataa agcattatca caaactattg agctattgga aggccacact 1920
tctctgatca gccagcacta ttatggttgc cttttccacg agcttctggc atttcagatt 1980

gctgaccgtc atccacagca agagagacaa cctgcggaag tgatacctca gcagatgata 2040
ggattttctg acccagctgt gtctacccta aaggctactg aatttttcgt tcctggttca 2100
gctgaatcca gcattattca tatggattta gatcgcagct tcctgccagt accttcagca 2160
gtgaatgcgt ctgtttttga aagttttgtc aggcagaaca tcactgattc tgaaaccgat 2220
gtccgttcca gcatccaaca gctggtgaaa gacagctatg gtttatccgc agctggttgt 2280
gcagaaatta tctacggaaa cacctccgta gcactcttca acaagctcgt cctttgttgc 2340
atgcaagaac agggcacctt gcttttcccc ttgggcacca atggccacta cgtctccgca 2400
gcaaagtttg tgaacgctag caccgtgact ataccgacaa atcccagttc agggttcagg 2460
atcgagccca aggttctagc tgacactctt aagaatgtat ctcggccgtg ggtgtatgtt 2520
tgcggcccca cgatcaaccc tactggcttt ctgtacagtg acagtgacat ccgagagctg 2580
ctctccgtat gcgctgaata tggagccagg gtagtgatag atacatcctt ctctggcctg 2640
gagtacgaga ccgatggatg gaggcagtgg aatctggcag gatgcctttc ctctttgaag 2700
cgttcagagc catcgttctc tgtggtcctg ctcggagagc tgtcctttgc gctgactgca 2760
ggtgggcatg actttgggtt tgtgatattg ggtgactcgt cgttggctga gacgttccat 2820
agtttttcta gcttgagtcg accgcacacc acattgaagt acactttcaa gaagctgctg 2880
gggcttaaga accagaagga ccagcatttc tcggatctga tagtggagca gaaggaggag 2940
ctgaagaatc gcgccaacca gttgatacag acactggaga gctgtggctg ggaagctgcc 3000
atcggttgcg gtggcatctc gatgctggcg aagcccaccg cgtacatggg gaaggccttc 3060
aaagctgctg gcttcgacgg cgagctggac gcgagcaaca tccgggaagc catcctgagg 3120
gccactgggc tgtgcataaa cagcagctcc tggacgggga tccctggcta ctgccggttc 3180
agctttgctc tggagagggg ggaattcgag cgtgcgatgg gatgcatagc tcggttcaag 3240
gagctggttc tgggaggcgc tcagatgaat ggtgcctgat ggagcaatcc atgttatacg 3300
taaccttggt gctgtgtagt agcgatgata ttgttccagg tgatggagat atcaaggaat 3360
aacaagagcc tacctgctgt gaagtcaatg tactactagg atggataggg cgcttcgatg 3420
cgccctttct gtcttcgaat tgctgtgctg tgcacggaga ataaacatat actattatat 3480
atatacatac aatttgcta 3499

<210> 6
<211> 1091
<212> PRT
<213> Zea mays

<400> 6
Met Ala Ala Leu Ala Gly Glu Asp Lys Asp Val Asp Ala Phe Leu Ala
1 5 10 15

Asp Cys Thr Ala Ser Gly Asp Ala Ala Tyr Gly Ala Ala Lys Ala Val
20 25 30

Leu Glu Arg Leu His Ala Pro Ala Thr Arg Pro Ala Ala Arg Arg Leu
35 40 45

Leu Gly Ala Val Arg Arg Arg Phe Ala Ala Ser Arg Ala Ala Gly Glu
50 55 60

Asp Cys Phe Arg Thr Phe His Phe Arg Ile His Asp Val Val Leu Asp
65 70 75 80

Pro His Val Gln Gly Phe Gln Gln Met Lys Lys Leu Thr Met Met Glu
85 90 95

Ile Pro Ser Ile Phe Ile Pro Glu Asp Trp Ser Phe Thr Phe Tyr Glu
100 105 110

Gly Leu Asn Arg His Pro Asp Ser Ile Phe Arg Asp Lys Thr Val Ala
115 120 125

Glu Leu Gly Cys Gly Asn Gly Trp Ile Ser Ile Ala Leu Ala Glu Lys
130 135 140

Trp Cys Pro Ser Lys Val Tyr Gly Leu Asp Ile Asn Pro Arg Ala Val

145 150 155 160

Lys Ile Ala Trp Ile Asn Leu Tyr Leu Asn Ala Leu Asp Asp Asp Gly
165 170 175

Leu Pro Ile Tyr Asp Gly Glu Gly Lys Thr Leu Leu Asp Arg Val Glu
180 185 190

Phe Tyr Glu Ser Asp Leu Leu Ser Tyr Cys Arg Asp Asn Lys Ile Glu
195 200 205

Leu Asp Arg Ile Val Gly Cys Ile Pro Gln Ile Leu Asn Pro Asn Pro
210 215 220

Glu Ala Met Ser Lys Ile Val Thr Glu Asn Ser Ser Glu Glu Phe Leu
225 230 235 240

Tyr Ala Leu Ser Asn Tyr Cys Ala Leu Gln Gly Phe Val Glu Asp Gln
245 250 255

Phe Gly Leu Gly Leu Ile Ala Arg Ala Val Glu Glu Gly Ile Ser Val
260 265 270

Ile Lys Pro Ser Gly Ile Met Val Phe Asn Met Gly Gly Arg Pro Gly
275 280 285

Gln Gly Val Cys Glu Arg Leu Phe Arg Arg Arg Gly Phe Arg Ile Thr
290 295 300

Lys Leu Trp Gln Thr Lys Ile Met Gln Xaa Ala Asp Thr Asp Ile Ser
305 310 315 320

Ala Leu Val Glu Xaa Glu Lys Asn Ser Arg His Arg Phe Glu Phe Phe
325 330 335

Met Asp Leu Val Gly Xaa Gln Pro Ile Cys Ala Arg Thr Ala Trp Ala
340 345 350

Tyr Met Lys Ser Gly Gly His Ile Ser His Ala Leu Ser Val Tyr Ser
355 360 365

Cys Gln Leu Arg Gln Pro Asn Gln Val Lys Lys Ile Phe Glu Phe Leu
370 375 380

Lys Asp Gly Phe His Glu Val Ser Ser Ser Leu Asp Leu Ser Phe Asp
385 390 395 400

Asp Asp Ser Val Ala Glu Glu Lys Ile Pro Phe Leu Ala Tyr Leu Ala
405 410 415

Ser Phe Leu Lys Glu Asn Lys Ser Asn Pro Cys Glu Pro Pro Ala Gly
420 425 430

Cys Leu Asn Phe Arg Lys Leu Val Ala Gly Phe Met Lys Ser Tyr His
435 440 445

His Ile Pro Leu Thr Pro Asp Asn Val Val Val Phe Pro Ser Arg Ser
450 455 460

Val Ala Ile Glu Asn Ala Leu Gln Leu Phe Ser Pro Ala Leu Ala Ile
465 470 475 480

Val Asp Glu His Leu Thr Arg His Leu Pro Lys Gln Trp Leu Thr Ser
485 490 495

Leu Ala Ile Glu Gly Arg Ala Asp Cys Asn His Ala Asp Gly Thr Val
500 505 510

Thr Val Ile Glu Ala Pro Arg Gln Ser Asp Leu Leu Ile Glu Leu Ile
515 520 525

Arg Lys Leu Gln Pro Gln Val Val Val Thr Gly Met Ala Gln Phe Glu
530 535 540

Ala Ile Thr Ser Ala Ala Phe Glu Asn Leu Leu Asn Val Thr Lys Asp
545 550 555 560

Val Gly Ser Arg Leu Phe Leu Asp Ile Ser Glu His Leu Glu Leu Ser
565 570 575

Ser Leu Pro Ser Ser Asn Gly Val Leu Lys Tyr Leu Ala Gly Lys Thr
580 585 590

Leu Pro Ser His Ala Ala Ile Leu Cys Gly Leu Val Lys Asn Gln Val
595 600 605

Tyr Ser Asp Leu Glu Val Ala Phe Ala Ile Ser Glu Asp Ala Ala Val
610 615 620

Tyr Lys Ala Leu Ser Gln Thr Ile Glu Leu Leu Glu Gly His Thr Ser
625 630 635 640

Leu Ile Ser Gln His Tyr Tyr Gly Cys Leu Phe His Glu Leu Leu Ala
645 650 655

Phe Gln Ile Ala Asp Arg His Pro Gln Gln Glu Arg Gln Pro Ala Glu
660 665 670

Val Ile Pro Gln Gln Met Ile Gly Phe Ser Asp Pro Ala Val Ser Thr
675 680 685

Leu Lys Ala Thr Glu Phe Phe Val Pro Gly Ser Ala Glu Ser Ser Ile
690 695 700

Ile His Met Asp Leu Asp Arg Ser Phe Leu Pro Val Pro Ser Ala Val
705 710 715 720

Asn Ala Ser Val Phe Glu Ser Phe Val Arg Gln Asn Ile Thr Asp Ser
725 730 735

Glu Thr Asp Val Arg Ser Ser Ile Gln Gln Leu Val Lys Asp Ser Tyr
740 745 750

Gly Leu Ser Ala Ala Gly Cys Ala Glu Ile Ile Tyr Gly Asn Thr Ser
755 760 765

Val Ala Leu Phe Asn Lys Leu Val Leu Cys Cys Met Gln Glu Gln Gly
770 775 780

Thr Leu Leu Phe Pro Leu Gly Thr Asn Gly His Tyr Val Ser Ala Ala
785 790 795 800

Lys Phe Val Asn Ala Ser Thr Val Thr Ile Pro Thr Asn Pro Ser Ser
805 810 815

```
Gly Phe Arg Ile Glu Pro Lys Val Leu Ala Asp Thr Leu Lys Asn Val
            820                 825                 830

Ser Arg Pro Trp Val Tyr Val Cys Gly Pro Thr Ile Asn Pro Thr Gly
        835                 840                 845

Phe Leu Tyr Ser Asp Ser Asp Ile Arg Glu Leu Leu Ser Val Cys Ala
    850                 855                 860

Glu Tyr Gly Ala Arg Val Val Ile Asp Thr Ser Phe Ser Gly Leu Glu
865                 870                 875                 880

Tyr Glu Thr Asp Gly Trp Arg Gln Trp Asn Leu Ala Gly Cys Leu Ser
                885                 890                 895

Ser Leu Lys Arg Ser Glu Pro Ser Phe Ser Val Val Leu Leu Gly Glu
            900                 905                 910

Leu Ser Phe Ala Leu Thr Ala Gly Gly His Asp Phe Gly Phe Val Ile
        915                 920                 925

Leu Gly Asp Ser Ser Leu Ala Glu Thr Phe His Ser Phe Ser Ser Leu
    930                 935                 940

Ser Arg Pro His Thr Thr Leu Lys Tyr Thr Phe Lys Lys Leu Leu Gly
945                 950                 955                 960

Leu Lys Asn Gln Lys Asp Gln His Phe Ser Asp Leu Ile Val Glu Gln
                965                 970                 975

Lys Glu Glu Leu Lys Asn Arg Ala Asn Gln Leu Ile Gln Thr Leu Glu
            980                 985                 990

Ser Cys Gly Trp Glu Ala Ala Ile Gly Cys Gly Gly Ile Ser Met Leu
        995                1000                1005

Ala Lys Pro Thr Ala Tyr Met Gly Lys Ala Phe Lys Ala Ala Gly Phe
   1010                1015                1020

Asp Gly Glu Leu Asp Ala Ser Asn Ile Arg Glu Ala Ile Leu Arg Ala
1025                1030                1035                1040

Thr Gly Leu Cys Ile Asn Ser Ser Ser Trp Thr Gly Ile Pro Gly Tyr
                1045                1050                1055

Cys Arg Phe Ser Phe Ala Leu Glu Arg Gly Glu Phe Glu Arg Ala Met
            1060                1065                1070

Gly Cys Ile Ala Arg Phe Lys Glu Leu Val Leu Gly Gly Ala Gln Met
        1075                1080                1085

Asn Gly Ala
   1090
```

<210> 7
<211> 610
<212> DNA
<213> Nicotiana tabacum

<400> 7

```
cagggttttg tcgaggatca gtttggcttg gggcttattg caagggcagt tgaagaaggt 60
atttctgtca taaagccatt gggcattatg atcttcaaca tgggaggccg tcctgggcaa 120
ggtgtttgca aacggttatt tgagcgccgt ggtcttcgtg ttaacaagct ctggcaaact 180
aaaattcttc aggcagctga cactgatata tcagctctag ttgaaattga aaagagtagc 240
atgcaccggt ttgaattttt catgggactt gttggagatc agccaatatg tgctcgaaca 300
gcatgggctt atggcaaggc tggtggtcgt atctctcatg ctttatctgt gtacagctgt 360
caacttcgtc agccaaatca ggtcagaaag atatttgagt tcataaaaaa tggattccat 420
gatatcagta attctctgga tttgtcattt gaggatgatg cagtagcaga tgaaaagatc 480
cctttcctag cttatcttgc tagtgtgctt aaggagaact ctgttttccc atatgaatca 540
ccagcaggaa gtagatggtt ccgcaatctg attgctggct ttatgaaaac ataccaccat 600
tttcctctta                                                       610
```

<210> 8
<211> 203
<212> PRT
<213> Nicotiana tabacum

<400> 8

```
Gln Gly Phe Val Glu Asp Gln Phe Gly Leu Gly Leu Ile Ala Arg Ala
  1               5                  10                  15

Val Glu Glu Gly Ile Ser Val Ile Lys Pro Leu Gly Ile Met Ile Phe
             20                  25                  30

Asn Met Gly Gly Arg Pro Gly Gln Gly Val Cys Lys Arg Leu Phe Glu
         35                  40                  45

Arg Arg Gly Leu Arg Val Asn Lys Leu Trp Gln Thr Lys Ile Leu Gln
     50                  55                  60

Ala Ala Asp Thr Asp Ile Ser Ala Leu Val Glu Ile Glu Lys Ser Ser
 65                  70                  75                  80

Met His Arg Phe Glu Phe Phe Met Gly Leu Val Gly Asp Gln Pro Ile
                 85                  90                  95

Cys Ala Arg Thr Ala Trp Ala Tyr Gly Lys Ala Gly Gly Arg Ile Ser
            100                 105                 110

His Ala Leu Ser Val Tyr Ser Cys Gln Leu Arg Gln Pro Asn Gln Val
        115                 120                 125

Arg Lys Ile Phe Glu Phe Ile Lys Asn Gly Phe His Asp Ile Ser Asn
    130                 135                 140

Ser Leu Asp Leu Ser Phe Glu Asp Asp Ala Val Ala Asp Glu Lys Ile
145                 150                 155                 160

Pro Phe Leu Ala Tyr Leu Ala Ser Val Leu Lys Glu Asn Ser Val Phe
                165                 170                 175

Pro Tyr Glu Ser Pro Ala Gly Ser Arg Trp Phe Arg Asn Leu Ile Ala
            180                 185                 190

Gly Phe Met Lys Thr Tyr His His Phe Pro Leu
        195                 200
```

<210> 9
<211> 14
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 9
ataccaagca tctt 14

<210> 10
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 10
ggataagaca gt 12

<210> 11
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 11
tatgaatctg atct 14

<210> 12
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 12
tcagggtttt gt 12

<210> 13
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 13
ttcaacatgg gagg 14

<210> 14
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 14
gagttcttca tgg 13

<210> 15
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 15
catgctttgt ctgt 14

<210> 16
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 16
gatttatcct ttga 14

<210> 17
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 17
gctcaatttg aggc 14

<210> 18
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 18

ttgaaatatc ttgct 15

<210> 19
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 19
catgcagcta ta 12

<210> 20
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 20
gcaatggtta acatc 15

<210> 21
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 21
cagggttttg tcgagga 17

<210> 22
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 22
tctgtcataa agcc 14

<210> 23
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC

OLIGONUCLEOTIDE

<400> 23
ttcaacatgg gagg 14

<210> 24
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 24
aagctctggc aaac 14

<210> 25
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 25
agctgtcaac ttcg 14

<210> 26
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 26
Pro Glu Asp Trp Ser Phe Thr Phe Tyr Glu Gly
1 5 10

<210> 27
<211> 17
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 27
Asp Lys Thr Val Ala Glu Leu Gly Cys Gly Asn Gly Trp Ile Ser Ile
1 5 10 15

Ala

<210> 28
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 28
Lys Val Tyr Gly Leu Asp Ile Asn Pro Arg Ala Val Lys Ile
1 5 10

<210> 29
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 29
Trp Ile Asn Leu Tyr Leu Asn Ala
1 5

<210> 30
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 30
Lys Thr Leu Leu Asp Arg Val Glu Phe Tyr Glu Ser Asp Leu Leu Ser
1 5 10 15

Tyr Cys Arg Asp Asn
20

<210> 31
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 31
Arg Ile Val Gly Cys Ile Pro Gln Ile Leu Asn Pro Asn Pro
1 5 10

<210> 32
<211> 28

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 32
Leu Ser Asn Tyr Cys Ala Leu Gln Gly Phe Val Glu Asp Gln Phe Gly
1 5 10 15

Leu Gly Leu Ile Ala Arg Ala Val Glu Glu Gly Ile
20 25

<210> 33
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 33
Phe Asn Met Gly Gly Arg Pro Gly Gln Gly Val Cys
1 5 10

<210> 34
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 34
Asp Thr Asp Ile Ser Ala Leu Val Glu
1 5

<210> 35
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 35
Pro Ile Cys Ala Arg Thr Ala Trp Ala
1 5

<210> 36
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 36
Ile Ser His Ala Leu Ser Val Tyr Ser Cys Gln Leu Arg
1 5 10

<210> 37
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 37
Val Lys Lys Ile Phe Glu Phe Leu Lys
1 5

<210> 38
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 38
Glu Lys Ile Pro Phe Leu Ala Tyr Leu Ala
1 5 10

<210> 39
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 39
Thr Val Ile Glu Ala Pro Arg Gln Ser Asp Leu
1 5 10

<210> 40
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 40
Glu Leu Ser Ser Leu Pro Ser Ser

1 5

<210> 41
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 41
Tyr Tyr Gly Cys Leu Phe His Glu Leu Leu
1 5 10

<210> 42
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 42
Ala Leu Phe Asn Lys Leu Val Leu
1 5

<210> 43
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 43
Gly Pro Thr Ile Asn Pro Thr Gly
1 5

<210> 44
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 44
Leu Arg Ala Thr Gly Leu Cys Ile Asn
1 5

<210> 45
<211> 26
<212> PRT

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 45
Gln Gly Phe Val Glu Asp Gln Phe Gly Leu Gly Leu Ile Ala Arg Ala
1 5 10 15

Val Glu Glu Gly Ile Ser Val Ile Lys Pro
20 25

<210> 46
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 46
Phe Asn Met Gly Gly Arg Pro Gly Gln Gly Val Cys
1 5 10

<210> 47
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 47
Ala Asp Thr Asp Ile Ser Ala Leu Val Glu
1 5 10

<210> 48
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 48
Gln Pro Ile Cys Ala Arg Thr Ala Trp Ala Tyr
1 5 10

<210> 49
<211> 10
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 49
Ile Ser His Ala Leu Ser Val Tyr Ser Cys
1 5 10

<210> 50
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 50
Gln Leu Arg Gln Pro Asn Gln Val
1 5

<210> 51
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC PEPTIDE

<400> 51
Glu Lys Ile Pro Phe Leu Ala Tyr Leu Ala Ser
1 5 10

<210> 52
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 52
cagggtttyg tngargayca 20

<210> 53
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 53
ccagtaacca cnacytgngg 20

<210> 54
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (15)
<223> where n=inosine

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 54
aarttyttra aygcnaayat hatg 24

<210> 55
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<220>
<221> misc_feature
<222> (6)..(9)
<223> where n=inosine

<400> 55
ttraanccna cytcrgcytc 20

<210> 56
<211> 19
<212> PRT
<213> Wollastonia biflora

<400> 56
Lys Phe Leu Asn Ala Asn Ile Met Ser Ile Pro Thr Glu Ala Glu Val
1 5 10 15

Gly Phe Lys

<210> 57
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: SYNTHETIC OLIGONUCLEOTIDE

<400> 57
atcatgtcta tccctacaga 20